IMPLEMENTACIÓN DE PETREL STUDIO EN YPF

Ana Docampo, Rosana Sners, Fernanda Raggio, Gustavo Gimeno, Horacio Deamichez

YPF, S.A

Palabras Clave: ***data management, petrel, convencionales y no convencionales, studio***

Abstract: "Studio" Implementation in YPF

These are challenging times for oil companies, more and more data is available, more and more applications to use them, this is why data management is becoming so important now. Being able to integrate and centralize the information that is needed in the right time to make the right decisions, is the goal.

Integration of the reservoir model information to be used in petrel, was possible in YPF due to this new centralized, fully integrated and efficiently preserved environment called Studio, which was first implemented in the unconventional exploration team, where the ability to centrally integrate all the available information, and capture project knowledge in the context of daily workflows, allowed them to manage information from basin to prospect scale. After a successful implementation for this group, the project was spread throughout petrel users in the company.

INTRODUCCION

El *Data Management* es fundamental en los entornos empresariales actuales. La demanda interna y externa para gestionar los datos obtenidos durante la operación hace imperativo el disponer de una versión única y confiable.

La importancia del data management como herramienta de acceso rápido y seguro a los datos de E&P, apunta a incrementar la eficiencia y reducir los costos de E&P, donde es pilar fundamental que los interpretes tengan disponible la información necesaria para trabajar sin invertir tiempo en carga de datos.

Desde hace aproximadamente 4 años en ypf se comenzó a utilizar el software de petrel para el modelado geológico, que debido a sus características se utilizaba con una configuración local. Cada usuario tenía su(s) proyectos en el disco local de la máquina. Bajo esta configuración se hacía imposible la administración de datos centralizada, por lo cual cada usuario se veía forzado a invertir gran parte de su tiempo en cargar sus propios datos; esta configuración aunada al crecimiento de la cantidad de usuarios y proyectos en petrel, comenzó a generar inconvenientes:

- La misma información era cargada muchas veces (1 por cada proyecto de usuario)
- Muchas horas de interprete estaban destinadas a la carga de datos
- Diferentes versiones de los mismos datos de referencia, lo que generaba resultados erróneos o imprecisos.
- Difícil trabajar en equipo, compartiendo información
- Multiplicidad de datos
- Diferentes sistemas de coordenadas en proyectos de la misma área geográfica.
- El gerenciamiento de datos era prácticamente imposible
- Difícil mantener la información actualizada en todos los proyectos de usuario
- Se perdían proyectos por fallas de disco

Esto generaba resultados imprecisos, pérdida de tiempo, y dificultad en la comunicación interdisciplinaria, ante esta problemática, el desafío era implementar una solución para centralizar la administración de los datos en una base de datos única para petrel, de donde se pudieran generar los proyectos de usuario (sin pasar por exportación e importación de datos) y así todos contaran con la misma versión de los datos de referencia y que la misma tuviese control de calidad, y posteriormente puedan compartir resultados.
El propósito de este trabajo es mostrar como pudimos centralizar la información que se utiliza para el modelado geológico en YPF, a partir de esta configuración.

METODOLOGIA DE IMPLEMENTACIÓN

La solución a implementar debía cumplir varios requisitos para ser fácilmente adquirida por usuarios que estaban acostumbrados a trabajar con tiempos de respuesta de disco local, además de ser eficiente en la carga de datos y colaboración de los mismos.
Anteriormente se habían realizado diversos intentos de implementar discos de red dentro de la infraestructura existente para que se utilizaran como repositorio de los proyectos petrel. Estos intentos habían sido infructuosos, ya que la velocidad de respuesta de los proyectos de petrel desde estos discos no era la adecuada para trabajar en línea con la cantidad de usuarios que existían. Los tiempos de espera eran largos para procesos comunes de interpretación y si al momento de salvar el proyecto se presentaban problemas o alguna interrupción de red, el proyecto podía corromperse.
Por otro lado, para que los usuarios comenzaran a trabajar en un mismo ambiente era necesario una tarea de data management para generar un único proyecto por cuenca, con la información validada de datos raw e información de interpretación que pueda extraerse de los diversos proyectos de usuarios.
Por estas razones, la implementación de Studio se dividió en 2 fases:

FASE 1: INFRAESTRUCTURA:

En vista que no se contaba con un servidor que permitiera dar los tiempos de respuesta que se necesitaban, se realizó un benchmarking assesment con Schlumberger, para configurar cada Workstation de la mejor manera para ejecutar petrel eficientemente y usar todos los recursos de la máquina y de la red. Se llegó a un acuerdo con un proveedor de infraestructura NAS para que bajo la modalidad de “try & Buy” nos prestaran un equipo para hacer las pruebas de concepto y eventualmente comprarlo si las pruebas resultan exitosas. Una vez instalado y configurado el equipo se ejecutaron pruebas con 20 usuarios al mismo tiempo ejecutando el proyecto espacialmente creado para el benchmarking de manera de tratar de sobrecargar las máquinas y la red. El detalle de estas pruebas pueden ser consultadas en el anexo Nro 1.
Estas pruebas fueron exitosas, por lo cual se adquirió el servidor con el que se hicieron las pruebas.
Por otro lado se hizo la compra de un servidor para la BD de Petrel Studio.
Cuando contamos con estos 2 equipos, se configuró el Studio Knowledge Environment, que constaría de:

- Servidor BD Oracle petrel
- Servidor de Disco de Alta disponibilidad.

En el servidor de BD Oracle, se instaló Studio y se configuró la bd.

El servidor de alta disponibilidad debería contar con lo siguiente:
Un directorio por cada cuenca, y dentro de cada cuenca:

1. Directorio de sísmica compartida: Este directorio tiene como finalidad ser el repositorio de los segy o la sísmica realizada de todos los proyectos petrel de los usuarios, de esta manera no se duplica la información sísmica por cada proyecto de usuario; sino que se realizan links que apuntan a la sísmica en ese directorio.
2. Proyectos de referencia: Para almacenar proyectos de lectura, donde se guarda información que no sube a la Base de datos, por ejemplo modelos.
3. Directorio de índices: Los índices permiten saber qué información se encuentra cargada en la base, y encontrarla fácilmente. Se actualizan diariamente para mantener actualizada la catalogación de datos de la base.
4. Directorio de templates: Se generan por cuenca, proyectos templates donde hay información general que ya está predeterminada, por ejemplo el sistema de coordenadas, templates de visualización de datos, información cultural, etc.
5. Directorio de proyectos de usuario: Cada usuario tiene su directorio con su(s) proyecto(s) para trabajarlos directamente desde este disco, de esta manera se vela por la seguridad del proyecto: Seguridad de acceso, mediante control de usuario y contraseña al disco de red. Y seguridad de que no se pierde la interpretación del usuario mediante el backup diario.

FASE II. DATA MANAGEMENT

Es crítico al iniciar un proyecto, la toma de decisiones más acertadas y confiables. Si consideramos que nuestro objetivo es tener la mejor calidad de información para el usuario final, en el menor tiempo posible, se tenía que decidir entre revisar la información de todos los proyectos petrel ya construidos durante 4 años por aproximadamente 50 usuarios; o construir una base de datos desde cero con la información oficial desde las bases de datos corporativas.
La opción más eficiente en cuanto a tiempo y cantidad de la información a cargar en Studio fue la opción de tomar la información fuente desde cero, no desde cada proyecto de usuario. Esta construcción se realizó de la siguiente manera:

1. Carga del datos raw desde las bd corporativas.
2. Información adicional de pozos que no estaba en las bd corporativas: se cargó desde diversas fuentes (legajo del pozo, informes de laboratorio, etc).
3. Interpretación de los usuarios: Desde algunos proyectos petrel previamente definidos, y con la participación de usuarios responsables por tipo de datos, se cargó interpretación de usuarios que ya estuviese en fase final para poder ser compartida a otros usuarios. Ver fig. 1.

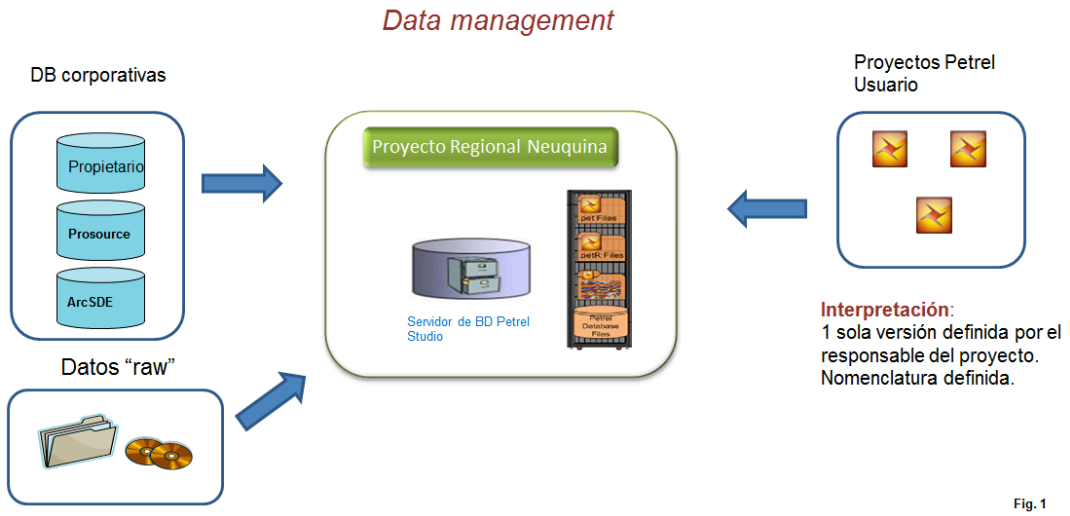


Fig. 1

En Studio los datos se manejan por entidades: Pozos, surveys 3D, líneas 2D. Una vez que las entidades son cargadas en Studio se genera el código único interno (GUID) que lo identifica a esa entidad como única para Studio y todos los proyectos petrel hacia donde se transfiera ese dato. A partir del momento en que estas entidades existen en la base, se pueden generar los proyectos de usuario para que estos puedan comenzar a trabajar.

Una vez que los datos de referencia estaban completos: pozos (desviaciones, logs, checkshots) y sísmica se comenzó a utilizar la base de studio como fuente para la generación de 5 proyectos petrel para que los usuarios de Exploración Convencional pudieran comenzar a trabajar, paralelamente se continuaba la carga de información adicional por pozo (geoquímica, punzados, petrofísica, presiones, coronas, logs de imagen, etc), de esta manera no hubo que esperar hasta finalizada la carga total de información para comenzar a usar la bd de Studio.

Con la implementación de petrel studio se implementa también un nuevo concepto para trabajar con petrel. Anteriormente se copiaban proyectos de un usuario a otro y se continuaban cargando datos de diferentes fuentes, sin control de calidad. Ahora el proyecto de petrel nace a partir de un ambiente controlado, con los mismos datos de referencia para todos los usuarios. Ver Fig. 2

Fig. 2

La información cargada en Studio abarca toda la escala de información desde lo macro a nivel cuenca hasta lo micro a nivel poro.

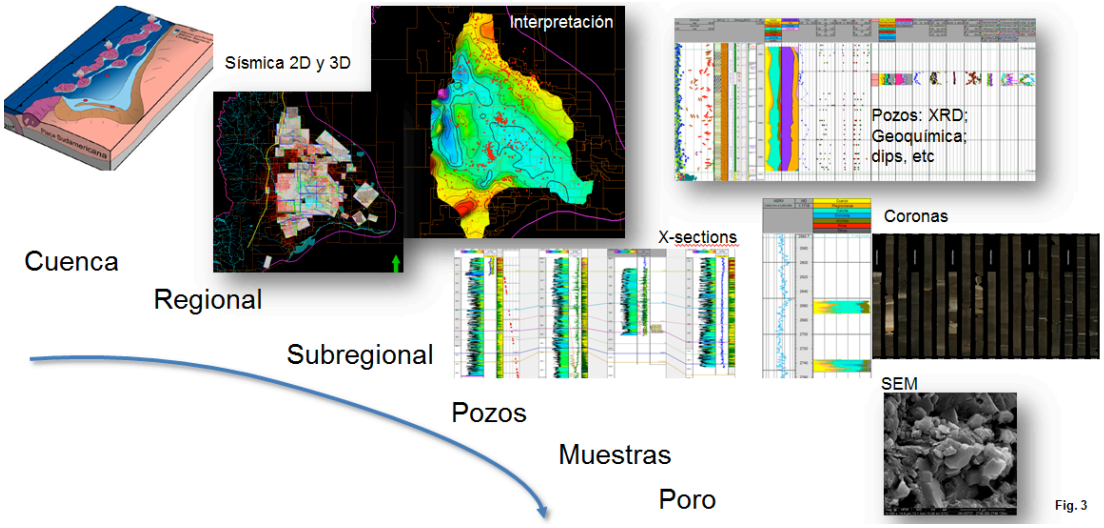


Fig. 3

COLABORACIÓN

Un equipo típico de Upstream consiste de Geólogos, geofísicos, ingenieros de reservorios y otros, en ocasiones, estos están ubicados en la misma oficina y en otras pueden estar separados físicamente. Para que un equipo multidisciplinario pueda funcionar efectiva y eficientemente, la base de datos y el software deben estar integrados.

Para YPF el reto es tener la información ordenada y centralizada para poder luego accederla desde la locación donde se encuentra el usuario. Esta es una fase en desarrollo aun, pero se tiene pensada una estructura como la que se ve en la figura 4

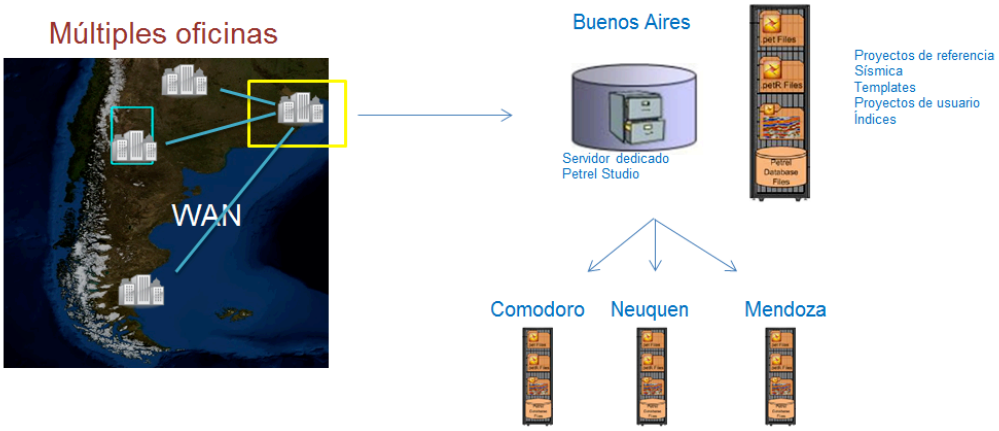


Fig. 4

Esta es una configuración semi-centralizada, porque si bien la BD está centralizada en Buenos Aires, el resto de la información que conforma el ambiente Studio esta regionalmente en cada servidor NAS. Estos servidores pueden sincronizarse una vez al día o una vez a la semana, y cada vez que el data manager cargue un nuevo dato sísmico.

De esta manera los usuarios que se encuentren en las áreas remotas pueden tomar la información de pozos directamente desde la BD en Buenos Aires y en cuanto a la información sísmica se toma el GUID desde la base de datos y el link al dato sísmico se realiza en cada servidor NAS local para mayor eficiencia.

La información de proyectos de usuario va a diferir en cada locación y este directorio puede o no estar sincronizado, con la dificultad de que la información iría en ambos sentidos, por lo cual, para compartir interpretación entre usuarios, sería más conveniente idear otra solución, como por ejemplo subir las interpretaciones “en progreso” o que aún no tienen versión final a un repositorio intermedio usado solamente para compartir información. Pero esta idea aún no ha sido probada en un piloto dentro de la compañía.

RESULTADOS

- Ahora es posible gerenciar los datos que se utilizan en petrel.
- Existe una sola versión (con control de calidad) de los datos de referencia.
- Un único sistema de coordenadas para la cuenca
- Los datos son cargados una sola vez
- La información es actualizada una sola vez consistentemente para todos los usuarios
- La información es indexada. Más fácil de buscar y más fácil de obtener.
- Más información puede ser cargada ahora, por lo cual hay más información disponible para interpretación
- Generación de nuevos proyectos en minutos
- Backup diario de la base y de los proyectos de usuario
- No hay duplicación de sísmica.

CONCLUSIONES

- Lleva tiempo construir una base de datos de conocimiento pero el esfuerzo se ve recompensado, por la mejor utilización de los recursos (personas y recursos informáticos)
- Se puede llevar una colaboración multidisciplinaria con las áreas de trabajo, aunque estén separados geográficamente, siempre que haya conexión de red.
- Desde el punto de vista del data management, se pone a disposición del usuario la información en tiempo y forma, pudiendo dedicar los recursos a la incorporación de más información con mayor control de calidad.
- Para el grupo de geocientistas la utilización de los datos para petrel provenientes de Studio, en vez de la generación masiva de proyectos "standalone" también trajo beneficios importantes
    - La información está organizada y clasificada, disponible en un solo lugar.
    - El equipo interdisciplinario puede trabajar integradamente, al acceder a la misma información de referencia.
    - Los tiempos de interpretación se optimizaron considerablemente, sobretodo en la fase inicial del proyecto donde los interpretes le dedicaban meses a la carga y validación de los datos de entrada.
    - Preservar el conocimiento. La idea de que haya un flujo que retroalimente la base de datos de Studio, permite que la interpretación final quede resguardada en el mismo lugar para que los futuros proyectos que se generen ya tengan ese avance en su plan de trabajo.

BIBLIOGRAFÍA

Petrotecnia, 2002. El Data Management y las empresas. Biblioteca IAPG, P. 8-13. Buenos Aires, Argentina.

Remasa, 2013. IAPG Data Management Workshop. Biblioteca IAPG, P. 8-13. Buenos Aires, Argentina.

ANEXO I

OPTIMIZACION DE LA PLATAFORMA TECNOLOGICA (IT)

El desafío de IT era conformar una solución que permita centralizar la información de los usuarios distribuida en los discos locales de cada estación de trabajo manteniendo el nivel de accesibilidad y performance que esto representaba, con el objetivo de lograr un gobierno completo de la gran cantidad de información (Big-Data) utilizada para el modelado geológico en YPF.
Acompañando las necesidades de Data Management en E&P se desarrolló una solución que permitiera desde Infraestructura ofrecer una plataforma de soporte para herramientas de acceso rápido y seguro a los datos aportando características de valor:

- ✓ Disponer infraestructura de alta disponibilidad y performance.
- ✓ Ser capaces de escalar acompañando el crecimiento del negocio.
- ✓ Asegurar la información permitiendo acceso granular a los archivos.
- ✓ Ofrecer alta tolerancia a fallas.
- ✓ Desarrollar funcionalidades avanzadas de protección y servicio.
- ✓ Implementar mecanismos de resguardo y backup de información.
- ✓ User friendly

PLATAFORMAS EVALUADAS

Con el objetivo de realizar una evaluación comparativa de las plataformas, que pudieran cubrir los requerimientos antes descriptos, se analizaron las siguientes infraestructuras:

- Cabezal NAS (Network Attached Storage)
- Servidor de Archivos

NAS:

Como sus siglas lo indican, un almacenamiento conectado a la red que se caracteriza por disponer diferentes tipos de file systems consolidados en una única plataforma escalar multidimensional. Así mismo ofrece: beneficios económicos, conectividad mejorada, altos niveles sostenidos de operaciones input/output por segundo (IOPS), eficiencia en almacenamiento, protección de datos, compatibilidad y durabilidad. Todas estas características soportadas por hardware y software de propósitos específicos.
La infraestructura NAS responde a una serie de capacidades críticas que los entornos empresariales actuales deben considerar cuando requieren desplegar arquitecturas de almacenamiento de archivos escalar.

Capacidad: Indica la eficacia de la plataforma para soportar grandes niveles de crecimiento lineal. Esto tiene en cuenta las limitaciones de capacidad escalar teórico y práctico, como ser capacidad máxima de archivos, cantidad de nodos y discos soportados por un file system, volúmenes o namespace.

Flexible: Se refiere a la capacidad de ofrecer distintos tipos de protocolos de acceso para ser consumidos por diferentes plataformas, entre ellas encontramos: NFS, CIFS, FTP, HTTP, etc.

Eficiente: Integrándose con capacidades específicas del storage, los sistemas NAS permiten el uso de funciones tales como: compresión, deduplicación, thin provitioning y automated tiering para reducir el TCO.

Compatible: Soporte a aplicaciones de terceros (ISV, independent software vendor), APIs de nubes públicas e hypervisores de mercado.

Administrable: Provee herramientas de automatización, management, monitoreo y reportes específicas para las funciones soportadas. Estas herramientas y programas están generalmente incluidas en una consola única de administración diseñada para mejorar la experiencia de uso mediante dashboards unificados con información de uso del sistema, alarmas y monitoreo en general.

Performante: Como pilar fundamental, estas plataformas ofrecen capacidad de crecimiento modular permitiendo incrementar sus niveles de operación en IOPS y ancho de banda que pueden ser entregados por el cluster según especificaciones máximas.

Resiliente: Se refiere a las capacidades ofrecidas por la plataforma para proveer altos niveles de disponibilidad. Dentro de las opciones ofrecidas encontramos: Alta tolerancia a fallas ocasionadas por roturas simultaneas de discos y/o nodos, técnicas de mitigación de fallas, protección nativa contra corrupción de datos y otras técnicas de resguardo (como snapshoots, réplica local y con distribución geográfica) que permiten minimizar los tiempos de RPO (recovery point objetives) y RTOs (recovery time objetives).

Seguro: Enfocado a proveer entornos seguros, los sistemas NAS disponen de funcionalidades nativas de seguridad para ofrecer control de acceso granular, encripción de la información bajo demanda, protección anti malware y configuración WORM selectivo.

ILM (information lifecycle management): Se refiere a la capacidad de manejo sobre el ciclo de vida de la información para el archivado y backup por tiempo prolongado, esta funcionalidad a su vez permite una optimización completa de la infraestructura.

SERVIDOR DE ARCHIVOS:

Comúnmente utilizado por usuarios como repositorio de diferentes tipos de datos no estructurados, estos servidores suelen ofrecer funcionalidades integradas sobre el sistema operativo consolidando diferentes tipos de servicios tales como: servidor de impresión, autenticación de usuarios, resolución de nombres de dominio, DHCP.

Si bien son muy útiles para varios aspectos básicos de la operación, cumpliendo con el objetivo de repositorio, debido a que en la mayoría de los casos carecen de funcionalidades de almacenamiento específico y escalable o las mismas son

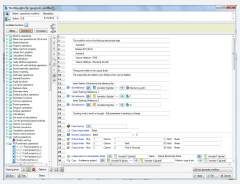
limitadas queda en desventaja frente a plataformas de propósitos específicos como la descripta anteriormente.

METODOLOGIA DE EVALUACION (Benchmarking Assessment)

Con el objetivo de realizar una evaluación precisa sobre las diferentes plataformas analizadas se definió un proyecto representativo, basado en un cubo sísmico de YPF y 1300 pozos con 3 sets de perfiles generados con algoritmos, sobre el cual se corrieron pruebas de performance y stress enfocados en tres workflows específicos.

Datos Geofísicos – Write

Geophysics Workflow



Lineas código: 114

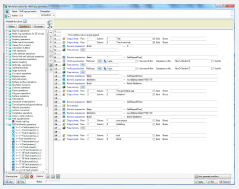
- Formato del Reporte Obtenido:

```
Task            Time In Seconds
Track 1                xx.xx
ZGY Generation         xx.xx
Track 2                xx.xx
Make horizon   xx.xx
RMS Amplitude  xx.xx
Gradient Magnitude xx.xx
                       ************
Total Time     xx.xx
```

Workflow automático 1 : Orientado a datos Geofísicos con foco en performance de escritura

Lecturas/Escrituras – I/O

Well Logs Generation Workflow



Lineas código: 30

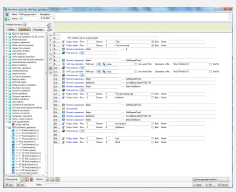
- Formato del Reporte Obtenido:

```
-------------------------------------------------
Task            Time In Seconds
generating logs        xx.xx
save project   xx.xx
                       *************
Total Time     xx.xx
```

Workflow automático 2: Utilizando datos de pozos y perfiles se genera gran cantidad de files, con foco en la medición de I/O.

Lecturas Random – Read

Prefetch to Cache Workflow



Lineas código: 13

- Formato del Reporte Obtenido:

```
-----------------------------------------
Task    Time In Seconds
Prefetch time...       xx.xx
```

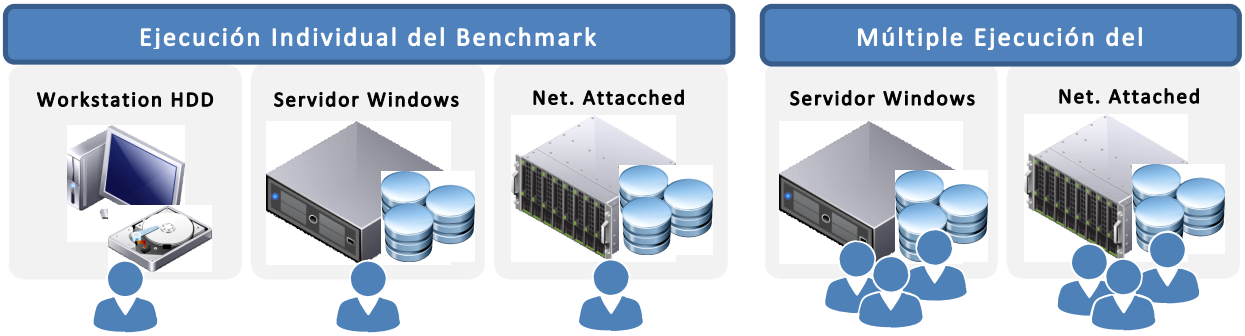
Workflow automático 3: Orientado a procesar gran cantidad de datos en memoria con foco en performance de lectura.

ESCENARIOS DE PRUEBA

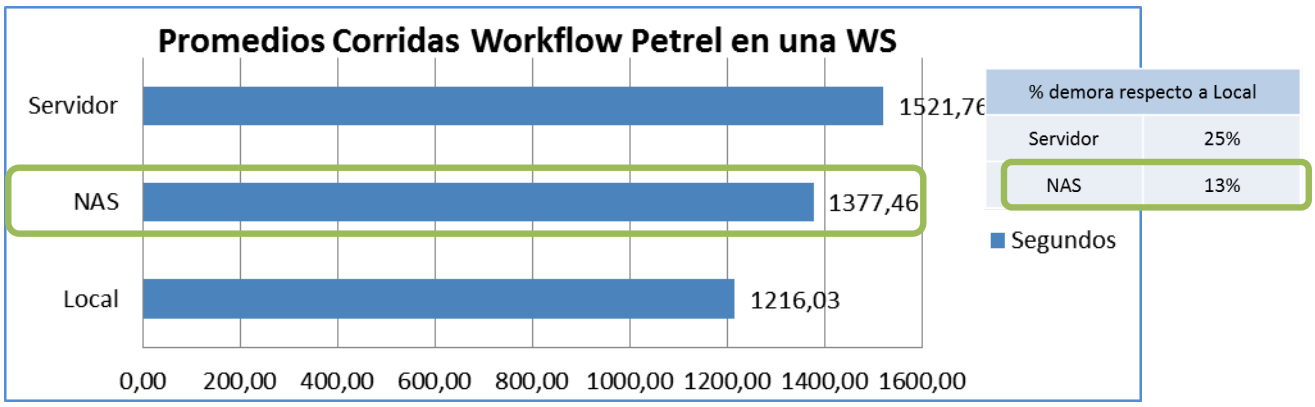
Inicialmente se realizó una verificación y adecuación de la conectividad y configuración de las Workstations asegurando un óptimo desempeño para las pruebas, esto incluyó un tunning en diferentes componentes. Cabe destacar la topología de la red distribuída, donde los usuarios se encuentran en el edificio sede de la compañía y el datacenter distante 1,5Km conectado por dos links de 1Gbps en L2 y dos links 10Gbps en L3.

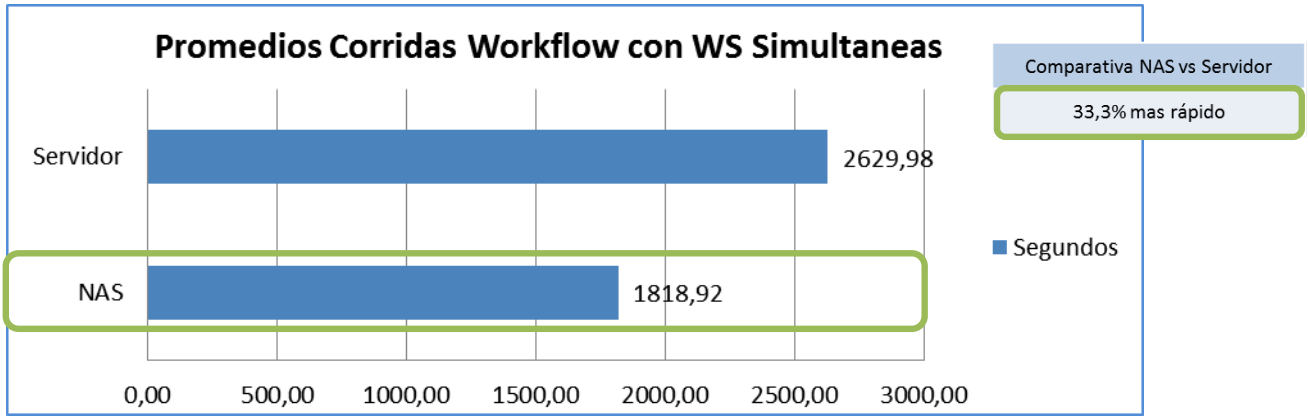
Se plantearon tres escenarios de prueba donde se corrieron los workflows automáticos antes descriptos en etapas con dos modalidades de ejecución:

- **Corridas individuales**: Se ejecutan los workflows en forma individual, desde un puesto de trabajo, y comparan los resultados.
- **Corridas múltiples**: Se ejecutan los workflows desde múltiples puestos de trabajo generando carga sobre los repositorios centralizados, Servidor de archivos y NAS.



RESULTADOS DE PRUEBAS





En los gráficos se puede observar las diferencias significativas de performance entre el file system centralizado provisto por un servidor de archivos tradicional (CIFS share) y un sistema NAS específico (CIFS share) para tal necesidad.

Destacamos que dada la concepción del sistema NAS, a mayor cantidad de usuarios concurrentes se incrementarán las ventajas de performance respecto a la ofrecida por un servidor tradicional.

CONCLUSIONES

En la búsqueda por construir una solución de infraestructura escalar y económicamente efectiva para el almacenado de grandes volúmenes de información, estructurada y no estructurada, que posibilite el correspondiente analizado y procesado de los datos, siendo este un factor clave dentro de los desafíos afrontados hoy en día en el ambiente IT, creemos conveniente en base a los resultados de las pruebas y al conocimiento adquirido, acompañar las necesidades de Data Management impulsadas por E&P con plataformas de propósitos específicos tipo NAS.

Estas plataformas se destacan como mejor alternativa frente a otras soluciones ofreciendo altos niveles de escalabilidad, flexibilidad, robustez y técnicas de protección sumadas a funcionalidades específicas para la optimización de datos que conforman un repositorio centralizado acorde a las necesidades del negocio.