

**La nube de datos  
ofrece a los CDO  
la ventaja que  
necesitan**



# Contenido

<b>Introducción: La aparición de datos alternativos</b>	03	<b>Una visión de futuro de la nube de datos</b>	17
Encontrar aristas en las finanzas	04		
<b>Una plataforma para poner en marcha sus datos</b>	05	<b>El papel del CDO ante la tendencia de la nube de datos</b>	18
Encontrar aristas en las finanzas	07		
<b>Una arquitectura moderna que desentraña la complejidad de los datos</b>	09	<b>Fuente de Snowflake</b>	19
Los retos de poner en marcha proyectos de datos	12		
<b>Casos de uso para los CDO</b>	13	<b>Acerca de PowerData</b>	20
<b>En qué hay que fijarse: Capacidades clave de la solución</b>	15		



# Introducción: La aparición de datos alternativos

**La búsqueda de una ventaja rentable es tan antigua como los propios mercados financieros. Y no hay mayor ventaja que la información propia y valiosa, unida a la capacidad de actuar antes de que el mercado le ponga precio. En los sectores actuales, esa ventaja se busca cada vez más a través de datos alternativos.**

Los conjuntos de datos alternativos ofrecen un sinfín de nuevas formas de encontrar una ventaja. Entre ellos se encuentran las tendencias de las redes sociales y los registros de transacciones, los datos de empresas privadas y las imágenes por satélite, los raspados de sitios web públicos y las etiquetas de geolocalización. Más de la mitad (53%) de los gestores de fondos de cobertura utilizan actualmente datos alternativos, y esta cifra va en aumento.<sup>1</sup> Entre los fondos de cobertura individuales que utilizaban datos alternativos en alguna medida el año pasado, más de un tercio gastó más de un millón de dólares en datos al año.<sup>2</sup>

Uno de los aspectos del valor de los datos proviene del volumen de los mismos: se dispone de más información detallada de una variedad de fuentes significativas que nunca antes. En 2020, cada persona en la Tierra creará 1,7 MB de datos por segundo, lo que supone 2,5 quintillones de bytes de datos producidos por los humanos cada día. Sólo en los dos últimos años se ha creado el 90% de los datos del mundo.<sup>3</sup>

Pero la identificación y adquisición de conjuntos de datos es sólo el comienzo de la estrategia de datos de una empresa de inversión en particular, o de cualquier empresa que en general quiera obtener valor añadido a través de los datos. La clave será la capacidad de integrar una amplia gama de conjuntos de datos personalizados, de compartirlos de forma flexible y de extraer las ideas clave a tiempo. Para ello, los conjuntos de datos alternativos residirán en la nube, donde se puede custodiar, acceder y compartir de forma mucho más rápida y económica que en cualquier otro lugar

Las enormes cantidades de datos que se producen requieren un cambio en la forma en que las empresas adquieren, procesan y utilizan esos datos para monetizarlos (interna o externamente) en diferentes casos de uso empresarial. Al igual que ocurre con la búsqueda de oportunidades en los mercados en desarrollo, las empresas que aprovechen el intercambio de datos moderno desde el principio obtendrán una ventaja a medida que la demanda de datos significativos siga creciendo y madurando.

1- "More Than Half of Hedge Funds Use Alternative Data to Gain Competitive Advantage" MondoVisione, May 5, 2020.

2- "How Much Are Managers Paying for Data?" Institutional Investor, July 9, 2020

3- "How Much Data Is Created Every Day in 2020?" TechJury, September 10, 2020.



# Encontrar bordes en las finanzas

La búsqueda de una ventaja rentable es tan antigua como los propios mercados financieros. Y no hay mayor ventaja que la información propia y valiosa, unida a la capacidad de actuar antes de que el mercado le ponga precio. En los sectores actuales, esa ventaja se busca cada vez más a través de datos alternativos.

El éxito ha sido muy lucrativo para pioneros cuantitativos como Renaissance Technologies, que gestiona activos por valor de casi 110.000 millones de dólares.<sup>4</sup> Pero, dado que la negociación algorítmica y los modelos de aprendizaje automático representan ahora un 80% del volumen total de operaciones en EE.UU., las máquinas están negociando cada vez más contra las máquinas.<sup>5</sup> Y eso significa que los numerosos imitadores de Renaissance tendrán más dificultades para hacer que la negociación cuantitativa sea rentable.

El enorme volumen de información que analizan muchos de estos proveedores de datos ilustra uno de los principales retos de trabajar con datos alternativos: adquirirlos es sólo el principio. Este complejo ciclo incluye procesos heredados, como

la transferencia de datos, la generación de informes, la distribución y, a continuación, volver a empezar cuando hay nuevos datos disponibles. Con más de 350 proveedores que venden un volumen de datos cada vez mayor, las empresas también deben evaluar la propuesta de valor de cada fuente potencial.

A medida que aumenta el número de compradores y vendedores, también lo hace la dificultad de mantener e integrar múltiples conjuntos de datos, fuentes y plataformas.

El problema de los bordes es que inevitablemente se erosionan. En el caso de los conjuntos de datos alternativos para los mercados financieros, los más populares acabarán perdiendo gran parte de su valor a medida que los mercados absorban y pongan precio a sus señales. Pero los datos crecen continuamente, por lo que no faltarán alternativas.

Las empresas que participan activamente en la economía de intercambio de datos hoy en día ya tendrán las conexiones y la infraestructura necesarias para hacerse con esas alternativas primero.

## New Technology Trends Change How We Use Data

### Rise of the Cloud



Cloud gives us the ability to scale and centralize data

### Explosion of Data



IoT, mobile, and social open up new opportunities for insight

### Diversification of Analytics



Analytics is growing in importance, everywhere, and for everyone

4- "Renaissance Technologies" Wikipedia, accessed October 9, 2020.

5- "Algo Trading Dominates 80% Of Stock Market" Seeking Alpha, January 1, 2019.



# Una plataforma: para poner en marcha sus datos

Los recientes cambios en la tecnología están modificando la forma en que las empresas trabajan con los datos, lo que facilita que la empresa los convierta en un activo real.

## 1. El auge de la nube

A diferencia de las plataformas tradicionales que se construyen pensando en la escasez de recursos, la nube ofrece una capacidad ilimitada para almacenar y procesar datos. Esto nos da la oportunidad de ejecutar un número muy elevado de cargas de trabajo con un alto rendimiento dentro de una única plataforma.

## 2. La explosión de datos

Los datos están creciendo en varios aspectos, como el tamaño, el formato/la estructura y la variedad.

El análisis de una amplia y profunda selección de información puede proporcionar conocimientos significativos, pero primero tenemos que ser capaces de acceder a ella.

## 3. Diversificación de la analítica

Cuándo, dónde y cómo hacemos análisis se ha transformado completamente. Hoy en día, todo el mundo necesita analizar datos. Se está convirtiendo en una parte central de todos los trabajos. Necesitamos no sólo saber lo que está ocurriendo ahora, sino también ser capaces de predecir lo que puede ocurrir en el futuro para tomar las mejores decisiones posibles en nuevos escenarios mediante el uso de análisis avanzados.

## 4. La adopción y democratización de la IA

La inteligencia artificial (IA) se está convirtiendo en un valor innegable para empresas de todas las formas y tamaños. Si tiene tareas repetibles en su cadena de valor, lo más probable es que utilice una tecnología de IA<sup>6</sup>. Pero el entrenamiento de los sistemas de IA requiere tanto la disponibilidad y gestión de grandes cantidades de datos para su creación, como el uso de dichos datos para hacerlos progresivamente más éticos y explicables.

Para aprovechar al máximo todo este potencial, las organizaciones tienen que enfocar primero su manejo de los datos de una manera completamente diferente.

6- [How To Offer New Human+Machine Value Propositions With AI](#) Forbes, June 11, 2020.



## Aprovechar estas tendencias tiene un gran impacto en el negocio



Toma decisiones empresariales mejores y más rápidas

**En primer lugar**, es necesario que se concentre en todos sus datos. Centralizar los datos en una única plataforma es vital para tomar decisiones empresariales mejores y más rápidas. Al reducir el tiempo de espera de los datos, los usuarios obtienen todos los datos que necesitan, cuando los necesitan. Y esto se traduce en nuevas oportunidades de negocio.



Reducir el coste de la ampliación de la gestión y el análisis de datos

**En segundo lugar**, la facilidad de uso y la flexibilidad que conlleva trabajar con los datos de una forma nueva le permite reducir el coste global de cómo escalar la gestión de datos y los análisis. Cuando se crea un único lugar para todos los tipos de datos, se puede ofrecer mientras se planifica la escala, sin añadir la sobrecarga de construir y mantener una arquitectura tradicional.



Crear una gran experiencia de cliente con datos

**En tercer lugar**, cuando sea capaz de acceder mejor a todos los datos a un coste menor, podrá centrarse en ofrecer una gran experiencia al cliente. Puede aplicar lo aprendido para mejorar la forma en que su empresa interactúa con los clientes. También puede crear productos con mayor capacidad de personalización y calidad, aumentando directamente su valor percibido en las primeras etapas del embudo (conocimiento y consideración).

# Evolución hacia la nube de datos

**Las organizaciones no pueden utilizar la tecnología de ayer para resolver los problemas de datos de hoy, y definitivamente no los problemas de mañana.**

A lo largo del tiempo, muchas tecnologías han intentado aprovechar esta oportunidad, pero ninguna de ellas ha logrado satisfacer plenamente la facilidad de uso, la escala, la concurrencia de procesos y el rendimiento necesarios para soportar las cargas de trabajo actuales.

Durante décadas, los data warehouses empresariales locales hicieron un trabajo decente al ofrecer el rendimiento necesario para analizar datos estructurados. Al mismo tiempo, eran relativamente fáciles de usar (gracias a las bases de datos SQL), por lo que este enfoque creció en adopción. Pero a medida que la dinámica de los datos y el análisis de los datos y la analítica, se hizo difícil soportar cargas de trabajo concurrentes. Y no todos los usuarios podían acceder a los datos que necesitaban. **En algunos casos, los usuarios de la empresa tenían que esperar más de lo previsto para acceder, lo que reducía el tiempo disponible para la toma de decisiones óptimas.**

La primera generación de data warehouses en la nube ayudó hasta cierto punto.

Ofrecían elasticidad y mejor rendimiento de costes que los enfoques anteriores. Pero dado que eran simplemente versiones basadas en la nube de sus almacenes de datos locales, seguían estando limitados en su capacidad de manejar todos los datos para todos los usuarios.

Los usuarios no podían acceder de forma fiable a los datos que necesitaban. Incluso cuando podían acceder a ellos, el coste de no obtener los datos en un momento dado inhibía muchos casos de uso críticos que requieren la toma de decisiones en tiempo real.

Tanto los almacenes de datos empresariales como las soluciones de almacenes de datos en la nube de primera generación no pueden admitir todos los datos para todos los usuarios porque están diseñados para todos los sistemas de un solo clúster, antes de que hubiera docenas de cargas de trabajo que pudieran ejecutar.



Aunque trasladar data warehouse de la empresa a la nube facilitó la gestión, no resolvió fundamentalmente el problema del sistema de cluster único.

Hadoop y los almacenes de objetos de los lagos de datos resolvieron parte del problema al hacer posible el manejo de datos tanto estructurados como semiestructurados. Pero sólo abordaron parcialmente el problema de conseguir muchos usuarios en la plataforma.

Al mismo tiempo, estos enfoques son complejos, difíciles de usar y limitados en términos de rendimiento. Uno de los principales problemas era que Hadoop no estaba diseñado para SQL. Esto aumentaba significativamente las dificultades de adopción y ralentizaba el tiempo de obtención insights empresarial.

**Data Cloud ofrece la primera solución que combina lo mejor de los data warehouses empresariales locales, la nube y el big data en una única plataforma con escala, rendimiento y concurrencia.** Estas nuevas plataformas de datos también ayudan a optimizar la gobernanza de los datos. Incluyen los sistemas de trazabilidad y linaje de datos de Hadoop (seguimiento desde el principio del ciclo de vida de los datos hasta su final), de modo que la

gobernanza puede obtenerse fácilmente en combinación con SQL, al tiempo que ayuda a reducir los costes de administración. Estas plataformas también proporcionan una fácil integración con las herramientas de gobierno de datos que están evolucionando en las nubes públicas. Otra ventaja con respecto a Hadoop es la seguridad por diseño. La mayoría de estas

plataformas garantizan la seguridad tanto cuando se almacenan los datos como cuando viajan para ser consumidos. Con las soluciones de primera generación, diseñadas principalmente para el almacenamiento y el tratamiento de los datos en bruto, la seguridad tenía que llevarse a cabo como un paso adicional, lo que introduce capas de coste, complejidad y exposición al riesgo.

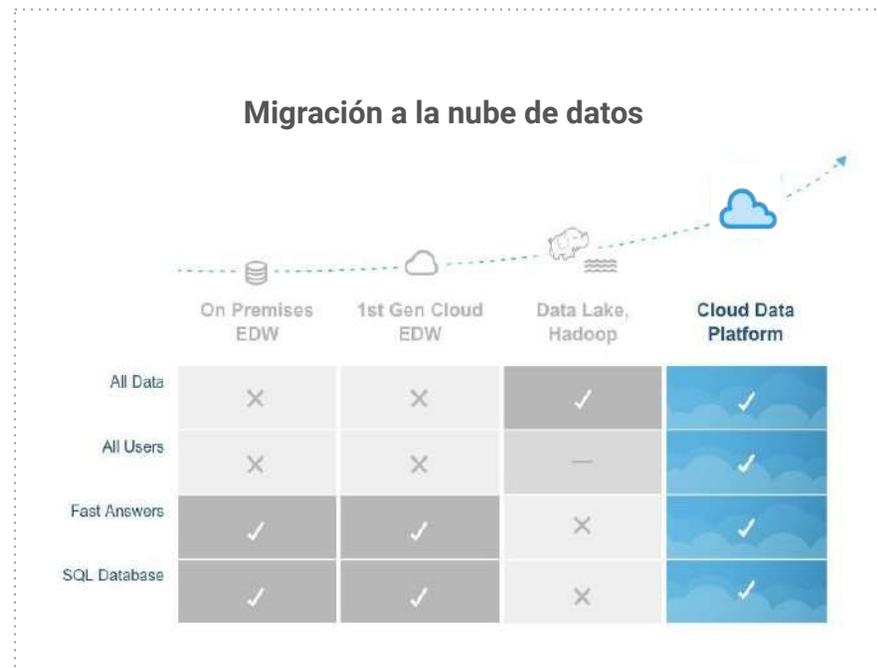


Figure 3: Journey to the Data Cloud

# Una arquitectura moderna que desentraña la complejidad de los datos

La mayoría de las organizaciones no pueden aprovechar plenamente el potencial de sus datos debido a las limitaciones de coste y complejidad de sus actuales arquitecturas de datos.

Las arquitecturas tradicionales suelen ser una intrincada red de tecnologías que se pegan para entregar los datos desde su fuente bruta hasta su destino. En cada paso de ese proceso (desde la integración de los datos hasta su transformación, agregación y entrega) se utilizan muchas herramientas diferentes, y cada punto de conexión entre herramientas dispares ofrece un posible punto de fallo. Los problemas de gobernanza y seguridad son también muy comunes en estos entornos.

Como resultado de lo anterior, suele ser necesario dedicar el 70-80% del **tiempo y el talento del equipo de datos a la administración y el mantenimiento de las capas inferiores de la infraestructura** (admisión y almacenamiento de datos). Sólo el 20-30% restante de los recursos disponibles puede dedicarse a ofrecer valor real a los usuarios de la empresa a través de enfoques como sistemas de informes, cuadros de mando, algoritmos y productos basados en datos.

Esta complejidad del flujo de trabajo actual, combinada con un volumen de datos en constante crecimiento, hace que los equipos técnicos tengan que estar más especializados para aportar valor en torno a los datos. También crea una mayor frustración para los equipos que trabajan en las capas finales de esta arquitectura (analistas, científicos de datos), así como para los usuarios de negocio, que a veces no comprenden los retos de escalar los problemas de muestras de datos limitados hasta los sistemas analíticos que procesan grandes volúmenes de datos.



## Arquitectura de datos tradicional

Complejo, costoso y limitado



Figure 4: Traditional data architecture

La Nube de Datos permite a las organizaciones construir una arquitectura de datos moderna que reduce la complejidad, el coste y las limitaciones presentes en otros enfoques.

### Una solución eficaz debe combinar

1. Almacenamiento centralizado para cantidades ilimitadas de datos estructurados, semiestructurados y no estructurados
2. Computación multi-clúster para ejecutar múltiples cargas de trabajo sin contención de recursos
3. Servicios en la nube para automatizar la seguridad de la administración y las tareas de la base de datos
4. Una capa agnóstica de la nube para ofrecer una experiencia consistente en todas las regiones y proveedores de la nube
5. Una interfaz única (a través de la web o de la consola) que proporciona una visión sencilla y centralizada del funcionamiento del sistema y del consumo de recursos en tiempo real para predecir posibles cuellos de botella y gestionar de forma óptima la plataforma

La arquitectura de Data Cloud admite la capacidad de ejecutar muchas cargas de trabajo diferentes con elasticidad, rendimiento y escala.

- **Data engineering** que agiliza la ingestión y la integración
- **Data Lakes** para almacenar grandes cantidades de datos en bruto
- **Data Warehousing** con rendimiento y concurrencia ilimitados
- **Data Science** para simplificar las cargas de trabajo e impulsar el rendimiento
- Carga de DevOps para construir **Data Applications**
- **Data Sharing** racionalizado entre consumidores y proveedores



Figure 5: The Data Cloud architecture

# Los retos de poner en marcha proyectos de datos

Aunque los proyectos que giran en torno a la explotación de los datos para la optimización y el desarrollo del negocio siguen siendo una prioridad para la mayoría de los ejecutivos, hasta el 85% de los proyectos de big data fracasan.<sup>7</sup> **Y del 15% de los proyectos que no fracasan, los directores generales afirman que sólo un 8% genera realmente valor, lo que se traduce en una tasa de éxito global del 2%.<sup>8</sup>** Junto con la explosión de datos en constante expansión, es difícil aumentar la tasa de éxito de la realización de proyectos de big data por dos razones:

La necesidad de reforzar la cultura de datos dentro de las empresas. Muchas organizaciones siguen luchando por contratar y retener a personal técnico cualificado con experiencia en datos.

Las expectativas internas a la hora de llevar a cabo proyectos de big data. Podría decirse que parte del problema es que los ejecutivos no saben cómo predecir el éxito de los proyectos de datos, ya que existen pocos enfoques estructurados para hacerlo.<sup>9</sup>

**Puedes tener todos los datos del mundo, pero si tu tecnología sólo puede ofrecer un prototipo una versión beta, o un volumen no escalable, en última instancia no se creará ningún valor real para su empresa.**

Afortunadamente, las nuevas herramientas diseñadas para la nube pueden ayudar a resolver los problemas de implementación de los proyectos de datos.

7- "Use This Framework to Predict the Success of Your Big Data Project" Harvard Business Review, February 25, 2020

8- "10 reasons why data science projects fail" Fast Data Science, May 8 2020.

9- "Use This Framework to Predict the Success of Your Big Data Project" Harvard Business Review, February 25, 2020.



# Casos de uso para los CDO

La Nube de Datos ofrece ventajas directas que pueden ayudar estratégicamente a los CDO en varios casos clave de uso de la gestión de datos.

## Caso de uso N° 1: Nuevas formas de monetizar los datos.

En el actual entorno empresarial altamente competitivo, la monetización de los datos es una poderosa herramienta en el arsenal de cualquier empresa. Data Cloud puede abrir nuevas oportunidades de negocio con conjuntos de datos "freemium". Esto significa ofrecer datos gratuitos que pueden ser útiles y valiosos para las empresas, al tiempo que se conservan columnas o capacidades adicionales para sus ofertas premium.

Un conjunto de datos agregados puede servir de referencia para las empresas que buscan líneas de base o datos comparables del sector. Además, diferentes mangas de productos de datos pueden proporcionar a las empresas valor sin necesidad de dar acceso a todo el conjunto de datos a todo el mundo. Todo esto le proporciona una mayor visibilidad y exposición a nuevos clientes potenciales. Si está empezando a considerar la monetización de los datos, seleccione un grupo de clientes para probar sus diferentes paquetes de datos y precios. Pueden ayudarle a comprender mejor el valor de sus datos y las consideraciones sobre el precio. sólo unos días al mes o al año, dejando ociosas las inversiones masivas el resto del tiempo. Busque una solución de plataforma de datos en la nube que cobre.

## Caso de uso N° 2: Control de costes

Puede ser una tarea frustrante entender cómo cobran los servicios algunos almacenes de datos en la nube. Para respaldar los presupuestos fijos y los planes de negocio, las organizaciones necesitan unos precios claros con respecto al funcionamiento de un almacén de datos en la nube. Los cargos ocultos, las cuotas y otras penalizaciones pueden ser frustrantes. Algunas soluciones de almacén de datos en la nube pueden cobrar en función de cuántos terabytes se escanean o cuántos terabytes se devuelven de una consulta. Además, es posible que no tenga la posibilidad de controlar los recursos informáticos disponibles porque se le asignan unidades de procesamiento o ranuras. Y si la consulta que ejecuta consume más que la asignación proporcionada, se le puede cobrar una prima por ese exceso.

Los almacenes de datos locales, en particular, hacen casi imposible reducir el hardware una vez instalado y provisionado. Esto es prohibitivo porque la mayoría de las implementaciones locales. La mayoría de las implementaciones locales están pre-dimensionadas y se compran para el período de máxima demanda, que puede ser de sólo unos pocos días al mes o al año, dejando las grandes inversiones ociosas el resto del tiempo. Busque una solución de nube de datos que cobre sólo por el tiempo y los recursos que utilice.



### Caso de uso N° 3: Seguridad

La seguridad debe ser fundamental para cualquier solución de datos. La seguridad integrada en el nivel de la solución significa que los clientes no tienen que desplegar y gestionar la seguridad ellos mismos. En su lugar, pueden centrarse en analizar sus datos en cualquier nube y región.

Una solución eficaz ofrecerá capacidades avanzadas de cifrado de datos. Todos los datos deben estar cifrados en movimiento, a través de Internet o de enlaces directos, y en reposo en los discos. También debe ofrecer autenticación de dos factores y de federación con inicio de sesión único. La autorización debe estar basada en roles, con la capacidad de establecer y aplicar políticas para limitar el acceso a direcciones de clientes predefinidas. Debe contar con certificaciones clave (como AWS y Azure) y admitir el cumplimiento de la privacidad de los datos con las normas reglamentarias.

### Caso de uso N° 4: Una plataforma para todos los datos de la empresa

Hoy en día se utilizan varios nombres para identificar dónde se encuentran los datos y cómo se utilizan, como almacén de datos operativos (ODS), fábrica de información corporativa (CIF), almacén de datos y mercado de datos. Cada término representa una forma diferente de agrupar los datos dentro de la empresa. Pero, por desgracia, estos diferentes grupos de datos también representan diferentes sistemas de datos.

Una solución de datos eficaz consolida las cargas de trabajo de los almacenes de datos, los lagos de datos, la ingeniería de datos, el intercambio de datos, la aplicación de datos y la ciencia de datos en una plataforma única y simplificada. Debe ofrecer una experiencia completa de gestión de datos empresariales que pueda proporcionar gobernanza, transacciones conformes con ACID, intercambio de datos en vivo, una huella global entre nubes y servicios totalmente gestionados.

### Caso de uso N° 5: Una plataforma para aprovechar los metadatos

Los sistemas de datos en la nube suelen incorporar de forma nativa mecanismos para poder analizar los metadatos generados por su uso.

Por ejemplo, respondiendo a las "tres Q":

**Qué:** los conjuntos de tablas (que engloban campos de información) que más se utilizan.

**Quién:** qué usuarios o aplicaciones consumen los datos almacenados.

**Cuándo:** en qué momentos se produce la demanda de información.



# En qué hay que fijarse: Capacidades clave de la solución

**En cualquiera de los casos de uso mencionados, los CDO deben buscar una serie de características y capacidades específicas que se combinen para ofrecer una solución de datos completamente formada.**

## Gestión de la concurrencia

Data Cloud puede procesar consultas y tareas en una fracción del tiempo que requieren las soluciones convencionales en las instalaciones. Cuando se combina con una escalabilidad lineal instantánea y casi infinita, las organizaciones pueden obtener el rendimiento exacto que necesitan, exactamente cuando lo necesitan. El tamaño del clúster informático puede seleccionarse en función de requisitos de rendimiento y se puede redimensionar en cualquier momento, incluso mientras se está ejecutando una carga de trabajo. A medida que las cargas de trabajo concurrentes aumentan, las capacidades de agrupación automática pueden añadir clusters de cálculo y distribuir las consultas entre ellos, eliminando la molestia de volver a agrupar los datos manualmente. Los clusters también pueden detenerse automáticamente cuando la carga de trabajo disminuye. Busque una solución que cobre sólo por los clusters activos, de modo que sólo pague por el rendimiento que utilice por segundo.

Dependiendo de cada caso de uso y del valor potencial que aporte a la organización, debería ser posible gestionar la escala necesaria de la concurrencia (ya sea mediante incrementos horizontales o verticales), aplicando al mismo tiempo las restricciones de seguridad adecuadas para el acceso a los datos y a la analítica. Este sistema de concurrencia también debe ofrecer una trazabilidad (en función de los nodos virtuales utilizados en la resolución) que permita al CDO tener una visión de 360 grados del caso de uso, pudiendo comparar el coste de computación medido con el beneficio generado por el mismo. Esto ayuda a simplificar el cálculo del retorno de la inversión (ROI).

## Clonación de copia cero

Esta función permite a los CDOs crear múltiples copias de tablas de datos, esquemas y bases de datos sin replicar los datos en sí. La clonación de copia cero permite la capacidad única de poner los datos a disposición de varios grupos de usuarios de forma casi instantánea sin replicar ningún dato.

En el caso de las empresas que tienen un único repositorio de datos que es utilizado por herramientas de inteligencia empresarial (BI) y de análisis, esta función puede simplificar las tareas de mantenimiento relacionadas con la alineación

de los datos contenidos en diferentes entornos. También puede poner rápidamente partes del repositorio a disposición de diferentes equipos de desarrollo, acelerando así el tiempo de entrega de las soluciones de información. Incluso puede realizar simulaciones sobre las posibles correcciones a realizar a nivel de tabla, tipo de datos o cálculo realizado y aplicarlas directamente desde el entorno clonado a la tabla de producción (si no se han producido más modificaciones desde la clonación).



### Integración

Una solución de datos eficaz debe admitir la integración con otras partes del ecosistema de datos y la arquitectura empresarial más amplia. La integración permite la flexibilidad de la carga de trabajo, lo que da a las organizaciones la libertad de trabajar con los datos en un lago de datos y construir tuberías de datos sólidas para agilizar la ingeniería de datos. La integración nativa con los principales lenguajes (como Python, R y Apache Spark™) ayuda a simplificar y acelerar las cargas de trabajo de la ciencia de datos.

### Gestión casi nula

Una verdadera plataforma de datos como servicio que se ejecuta en la nube elimina prácticamente la carga de la gestión de la infraestructura asociada a las plataformas tradicionales y las soluciones de big data. Las mejores soluciones se encargarán automáticamente de la infraestructura, la optimización, la disponibilidad y la protección de los datos, para que usted pueda centrar su tiempo y atención en el uso real de sus datos.

Esta característica debería ayudar al CDO y a su equipo a centrarse en la gobernanza del uso de los datos, así como en las medidas normativas y de seguridad que deben aplicarse.

Debería ayudar a reducir las complejas rutinas técnicas asociadas a la optimización de la plataforma para que esté preparada para satisfacer los picos de demanda de consumo de datos.

### Copia de seguridad y recuperación de datos

Los administradores de sistemas a veces necesitan capacidades de copia de seguridad de datos cuando se comete un error (por ejemplo, una tabla o un esquema se cae en producción) o una versión fallida requiere una reversión de la base de datos (por ejemplo, una nueva operación ETL corrompe los datos). Las organizaciones deben ser capaces de restaurar los datos basándose en un punto en el tiempo o un ID de consulta a nivel de base de datos, tabla y esquema. Algunas soluciones incluso ofrecen sólidos servicios de apoyo a la recuperación de datos que proporcionan una ventana de tiempo aún más amplia para restaurar los datos en caso de emergencia.

### Data Sharing

La capacidad de compartir datos de forma segura puede permitirle compartir sus datos con otros usuarios sin tener que copiar o transferir datos entre cuentas. Los propietarios de los datos pueden conceder un acceso seguro y curado a copias de datos de sólo lectura sin las cargas y complejidades de las herramientas de terceros como las tecnologías FTP y EDI.

Gracias a las nuevas funciones de intercambio de datos que posibilita este tipo de plataforma, se puede generar un nuevo ecosistema en el que los equipos de estrategia y monetización de datos pueden definir y planificar con el CDO. Esta colaboración puede ayudar a encontrar nuevas formas de generar valor a partir de los datos sin necesidad de recurrir a intermediarios innecesarios (siempre que la naturaleza de los datos que se comparten lo permita).

**Este tipo de paradigma de intercambio de datos que utiliza un marco tecnológico común entre los productores de datos existentes y los consumidores en el mercado también puede ayudar a diseñar nuevos recorridos de datos y modelos de negocio que antes eran impracticables debido a las limitaciones de información o de tiempo.**



# Una visión de futuro de la nube de datos

**Los datos se convertirán en uno de los mayores activos de cualquier empresa, si no lo son ya.**

Incluso cuando una empresa es consciente del problema, los silos pueden persistir, generando problemas colaterales y ralentizando las iniciativas empresariales. En primer lugar, es difícil unir, aumentar e integrar los datos de los diferentes silos, lo que impide a las organizaciones obtener una visión crítica de los conjuntos de datos. En segundo lugar, la fragmentación y proliferación de los datos crea una pesadilla de gobernanza de datos. **Y, por último, las organizaciones acaban pagando mucho más de lo que necesitan porque es imposible gestionar eficazmente el uso de la capacidad en un entorno fragmentado. La nube de datos elimina prácticamente todos estos problemas.**

Una solución de datos eficaz ayuda a las organizaciones a evitar que los datos queden atascados y en silos en la infraestructura y las nubes de aplicaciones, así como en los sistemas locales.

Una solución óptima le ofrece una única fuente de verdad para sus propios datos y la capacidad de compartir instantáneamente datos en vivo con otros de forma gobernada y segura. Al ser virtual y sin fricciones, no hay que copiar ni replicar los datos.

Las organizaciones designan los datos para compartirlos y les conceden permisos. El destinatario puede entonces procesar los datos en el lugar, sin obtener nunca la custodia física de los mismos. Esto significa que los datos están siempre actualizados, con latencia cero, y que cualquier actualización de los conjuntos de datos compartidos por los proveedores está disponible al instante para sus consumidores.

Este enfoque también puede trascender la nube pública y las fronteras geográficas. Las organizaciones pueden adquirir y compartir datos entre sí concediendo un acceso gobernado a conjuntos de datos específicos. Pueden estar ubicados en diferentes nubes y en diferentes regiones geográficas de una nube.

De vez en cuando, hay un cambio de nivel que realmente altera el juego. **La nube de datos** es exactamente eso: una oportunidad para movilizar completamente sus datos al servicio de su negocio. Proporciona un enfoque mucho más

agnóstico de la tecnología subyacente. **Ayuda a las empresas a centrarse en el uso interno y en la posible monetización a través de modelos de negocio nuevos u optimizados.**



# El rol del CDO ante la tendencia de la nube de datos

**El crecimiento de los datos sigue aumentando de forma exponencial, alcanzando una cifra estimada de 140 zettabytes en 2024.<sup>10</sup>**

**La arquitectura de la Nube de Datos simplifica tanto la ingestión como la gestión de los datos de forma más óptima que las arquitecturas de primera generación comentadas anteriormente.**

El reto inicial para un CDO es el diseño y evolución de la estrategia de datos dentro de su empresa. Los Data Cloud facilitan una estrategia de datos para trabajar con casos de uso de la nube híbrida, ya que permiten el despliegue en diferentes nubes públicas (de forma agnóstica y transparente) y aseguran la disponibilidad en el tiempo.

Otro de los retos a los que suelen enfrentarse los CDO es el diseño de data lakes que permitan de forma nativa la explotación y el autoservicio de la información ingerida. Gracias a la Nube de Datos, los lagos de datos pueden ser utilizados de forma segura sin necesidad de capas adicionales de tecnología que pueden hacer que estas arquitecturas sean más caras de diseñar y más

complejas en términos de gobernanza cuando las empresas necesitan escalar. Por otro lado, cada vez es más importante y necesario poder enriquecer los modelos analíticos con datos de fuera de la empresa, o monetizar parte de ellos (una vez anonimizados). En ambos casos, los CDO pueden pensar en la monetización de los datos eliminando ciertas barreras que existían antes de poder utilizar las funciones de "compartir datos".<sup>11</sup>

Aunque no depende exclusivamente del CDO, la "seguridad por diseño" es otra de las necesidades que las empresas tratan de incorporar. Dentro del ciclo de vida de los datos la seguridad de los datos es otro punto importante al que un CDO suele prestar atención cuando diseña una estrategia de datos.

Las nuevas plataformas de datos la incorporan en su ADN y permiten al CDO desplegar de forma transparente la securización progresiva de la explotación de los datos para cualquier equipo que utilice la plataforma.

Además de una adecuada gestión de la concurrencia, la Nube de Datos también puede permitir al CDO tener una trazabilidad exacta del coste de cada caso de uso, gracias a los metadatos almacenados en los recursos computacionales. Esto permite un cálculo exacto del VID así como un cálculo del consumo/coste de estos recursos.

Después de las reflexiones anteriores, podemos concluir que la Nube de Datos seguirá creciendo y se convertirá en una pieza estratégica dentro de las modernas pilas de arquitectura de datos (tanto físicas como virtuales). Estarán orientadas al pago por uso, para que las empresas puedan optimizar los recursos en su uso de los datos como activo y olvidarse de los procesos heredados del siglo pasado.

10- "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2024" Statista, February 5, 2021.

11- "Data Sharing Is a Business Necessity to Accelerate Digital Business" Gartner, December 3, 2020.



# Fuentes de Snowflake

Esta guía es una traducción al español de la versión original [“The Data Cloud Gives CDOs the Edge They Need”](#) publicado por Snowflake



PowerData, es una compañía multinacional de origen español con destacada presencia regional. Desde que se fundó en el año 2000 ha ayudado a más de 100 organizaciones de diversos sectores, categorías y nichos, quienes han depositado su confianza en el expertise y Know-how de PowerData para acelerar su camino hacia la transformación digital.

El alcance de la especialización en gestión de datos en la nube ha sido clave para que una empresa con ambiciones globales como Snowflake haya confiado en PowerData como partner para entregar lo mejor de su tecnología Data Cloud a las organizaciones.

Te invitamos a explorar en todos los proyectos donde aportamos valor con la gestión de datos en [powerdata.es](https://powerdata.es)

