

El Datawarehouse: El motor que impulsa el análisis

Por qué el core del análisis de datos es crítico en las decisiones que se toman basadas en los datos



Contenido

Por qué las organizaciones se están quedando atrás en la revolución analítica	03	El almacenamiento de datos creado para la nube como servicio	12
El camino hacia el análisis transformador está repleto de obstáculos	05	Las ventajas empresariales de implementar un datawarehouse en la nube	16
Las implicaciones empresariales de los obstáculos analíticos	08	Caso de éxito: de la implementación a la obtención de beneficios en menos de un año	17
Por qué el datawarehouse sigue siendo relevante	10	Acerca de PowerData	18
La nube es el lugar natural para el datawarehouse moderno	11		



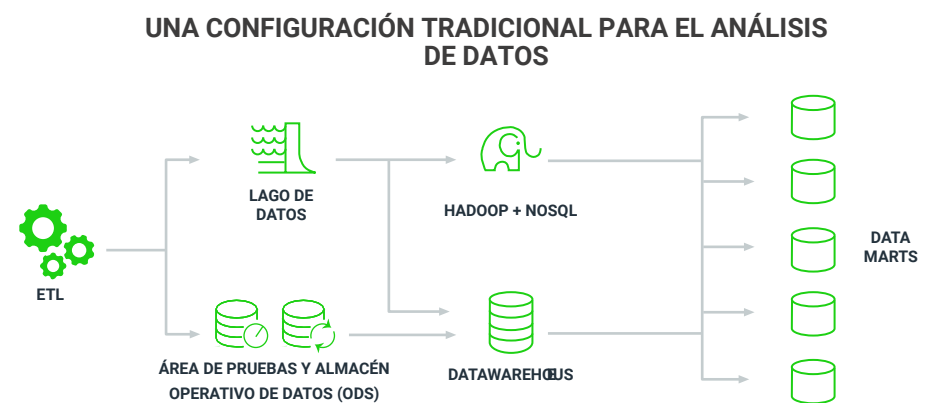
Por qué las organizaciones se están quedando atrás en la revolución analítica

Según McKinsey, el 86 % de los ejecutivos afirma que su organización solo ha sido eficaz moderadamente a la hora de lograr el principal objetivo de sus programas de análisis y datos. Y, lo que resulta aún más preocupante, más de un cuarto de estos profesionales afirma que la empresa ha sido ineficaz.

El hecho de que la mayoría de las iniciativas de datos no cumplan las necesidades reales de las empresas actuales o simplemente fracasen completamente al generar cualquier valor revela dos tendencias significativas. En primer lugar, las empresas están más centradas que nunca en el análisis de datos. En segundo lugar, es probable que la mayor parte de los fracasos nazcan en lo más profundo de la plataforma de análisis de datos de una organización.

Los productos de análisis de datos siguen superando la tecnología en la que se basan. Estos proveedores siguen ofreciendo soluciones avanzadas, pero intuitivas, que captan la atención de los profesionales de todos los niveles, departamentos y funciones empresariales, lo que representa la cara visible del servicio. En la cara oculta, tiene lugar el mismo proceso, donde un número de hábiles proveedores de ETL (por sus siglas en inglés, «extraer, transformar y cargar») y ELT (por sus siglas en inglés, «extraer, cargar y transformar») extraen los datos internos y externos de una empresa para introducirlos en un repositorio central con el fin de que las soluciones de análisis de datos realicen el trabajo.

Lo que queda entre medias es donde reside el problema. Este repositorio central suele adoptar la forma de un datawarehouse basado en tecnología heredada. Es decir, una empresa ha cambiado a una solución noSQL como, por ejemplo, Hadoop. Más recientemente, las organizaciones han reconocido las ventajas, y las muchas desventajas, de estas dos soluciones y han respondido con una alternativa híbrida que utiliza ambos sistemas en un tímido intento por mejorar su análisis de datos. Por tanto, las empresas se ven forzadas a añadir otro nivel de complejidad, los denominados «data marts», que se convierten en una necesidad analítica con este modelo, pero que constituyen uno de los grandes quebraderos de cabeza de la gestión de datos.



Por qué las organizaciones se están quedando atrás en la revolución analítica

Mientras tanto, las empresas están recopilando más datos que nunca. Y el volumen, la variedad y la velocidad de los datos indican que la mayoría de ellos se originan y se almacenan en la nube. Hay disponibles petabytes de datos, pero estos datos se quedan obsoletos en soluciones de almacenamiento ajenas a los datawarehouses tradicionales, ya que estos sistemas legacy (o heredados) no pueden procesar con eficacia nuevos tipos de datos semiestructurados. Es decir, resulta demasiado costoso almacenar una mayor cantidad de datos en sistemas tradicionales que no han sido creados específicamente para la nube. Estos sistemas, creados cuando la recopilación de datos era mucho menor y un número bastante inferior de usuarios necesitaba acceder a la información, siguen teniendo dificultades con su tecnología y su arquitectura heredadas.

Y, aunque Hadoop y otros sistemas noSQL han conseguido introducir diversos tipos de datos, obtener el mejor análisis de la información es prácticamente una tarea imposible. Estos sistemas han dejado a las empresas desprovistas de la base adecuada para convertir sus crecientes ansias de datos en verdadera información empresarial.

En esta guía electrónica, examinaremos los retos a los que se han enfrentado las empresas al convertir los datos en un conocimiento empresarial transformador. Además, obtendrá información sobre las tecnologías que están ahora disponibles para que las empresas sigan el ritmo de la revolución analítica. Gracias a un datawarehouse en la nube moderno, su organización y usted podrán convertirse en todo unos expertos.



El camino hacia el análisis transformador está repleto de obstáculos

Hasta ahora, se supone que su organización ha estado haciendo cosas transformadoras con el análisis. Sin embargo, los obstáculos tecnológicos o arquitectónicos creados por los datawarehouses heredados y los sistemas noSQL siguen limitando el análisis de datos necesario para mejorar constantemente la experiencia de los usuarios de los datos, agilizar las operaciones, hacer crecer la experiencia de los clientes de su empresa y permitir que su organización se convierta en líder del mercado y perdure como tal.

Estas frustraciones entran dentro de tres categorías principales:

- Obtener acceso a los datos
- Procesar el análisis de forma rápida y consistente
- Limitaciones de los Recursos Humanos

Obtener acceso a los datos

Analistas y usuarios de la empresa, a partes iguales, deben esperar a menudo horas, si no días, a que se carguen los datos y se ejecuten las consultas. Para aliviar el peso que tienen estas cargas de trabajo sobre las operaciones, con frecuencia debe esperar hasta altas horas de la noche para que los datos se carguen en sus datawarehouses. Sin embargo, para las empresas globales de hoy en día que trabajan las 24 horas del día, los datos de su almacén son siempre inferiores a los actuales. Y las nuevas formas de datos semiestructurados, como JSON, son sencillamente demasiado pesadas de cargar en un datawarehouse heredado. Por tanto, los datos permanecen en silos fuera del datawarehouse. Básicamente, los resultados de las consultas serán imprecisos e inadecuados, o estarán obsoletos, debido a la información limitada desactualizada que se incluye en el almacén.

Procesar el análisis de forma rápida y consistente

Los conjuntos de datos de su datawarehouse tienen una amplia aplicación en toda su empresa. Sin embargo, la ejecución de consultas simultáneas, las cargas de datos y los trabajos de desarrollo ralentizarán la velocidad de la mayoría de datawarehouses tradicionales a paso de tortuga, o aún peor. De hecho, puede que incluso

los administradores de bases de datos eliminen consultas que consideren menos importantes para que puedan ejecutarse trabajos más importantes.

Cuando esto sucede, los usuarios de los datos pueden desarrollar soluciones alternativas que a menudo no hacen más que agravar la situación.

A continuación, le mostramos algunas opciones habituales, junto con sus consecuencias, que suelen desarrollarse en un intento por combatir la latencia:

• Almacenamiento de datos en silos

Solución alternativa: el equipo de TI puede dejar los datos en su ubicación original, fuera del datawarehouse, para evitar esperar procesos ETL lentos. Consecuencia: los usuarios se ven limitados por resultados de consultas restringidos y la incapacidad de realizar análisis en todos los datos de una empresa desde un mismo repositorio. O, lo que es peor, intentan integrar los datos manualmente, lo que puede crear vistas inconsistentes. A menos que conozcan bien los datos, esto puede dar lugar a resultados incorrectos si los datos contenidos en estos



silos no se unen adecuadamente.

- **Usar «spreadmarts»**

Solución alternativa: los usuarios extraen grandes conjuntos de datos, posiblemente durante la noche, que colocan en hojas de cálculo de Excel para ejecutar sus análisis sin conexión.

Consecuencia: cuando esta práctica se generaliza, diferentes usuarios extraerán y guardarán distintas versiones de los mismos datos para desarrollar a continuación varias consultas analíticas y sacar diversas conclusiones. En esta situación, también resulta habitual que los usuarios integren otras fuentes de datos desgobernadas en los datos extraídos, sesgando aún más los resultados. En definitiva, ninguno de los datos se puede auditar.

Cada minuto que su equipo de análisis e inteligencia empresarial dedica a esperar los datos es un minuto que no pueden invertir en la obtención de información a partir de dichos datos. Mientras las empresas se basen en sistemas de datos que convierten el acceso a los mismos en un proceso lento y engorroso, los usuarios lograrán un menor volumen de trabajo analítico.

Seguirán buscando soluciones alternativas, lo que provocará un almacenamiento aún mayor de los datos en silos. Su empresa se quedará con una imagen incompleta o imprecisa de sus datos, mientras que los analistas no contarán con el tiempo necesario para encontrar información o ideas en los datos de los que disponen.

Limitaciones de los Recursos Humanos
A pesar de tanto hablar de la «democratización de los datos», las empresas siguen requiriendo que el equipo técnico ejecute sus datawarehouses heredados, físicos o en la nube. Los profesionales del análisis de datos y de TI, con todos los niveles de habilidades, siguen siendo el recurso humano más buscado en una organización. Modis, el proveedor global de personal de TI, estima un crecimiento del 12 % en los trabajos de TI para 2024, casi el doble del crecimiento previsto para el resto de industrias.

Con un datawarehouse tradicional, las empresas requieren que el personal:

- Cargue y transforme los datos en la estructura adecuada para una base de datos SQL.
- Gestione la base de datos y el hardware requeridos para mantenerla.

- Ayude a los usuarios de la empresa con sus datos de «autoservicio».

- Ayude a los usuarios de la empresa a desarrollar y ejecutar visualizaciones para un análisis más complejo.

- Escriba y ajuste complejas consultas SQL para llevar a cabo un análisis significativo de los datos.

El uso continuo de tecnología de datawarehouse heredada ha dado lugar a un problema adicional y más reciente relacionado con los recursos humanos. El aumento de datos semiestructurados de fuentes, tales como dispositivos móviles, plataformas de redes sociales y datos de máquinas, ha puesto de manifiesto las limitaciones del datawarehouse tradicional. Ante esta situación, muchas empresas han recurrido a sistemas noSQL como, por ejemplo, Hadoop, para crear un «lago de datos» (también conocido como «data lake»). Sin embargo, consultar estos lagos de datos es bastante más complicado que ejecutar una consulta SQL en un datawarehouse relacional y requiere diferentes conocimientos.

Por tanto, para Hadoop y otras soluciones noSQL, el problema de recursos humanos es



aún más grave. Estos sistemas solo han aparecido en los últimos 10 años y no han generado mágicamente los resultados que el revuelo inicial había previsto. Los profesionales con un alto grado de cualificación, y que son expertos en estas soluciones, representan una parte fundamental para que un proyecto Hadoop tenga un éxito moderado. Y, lo que es más importante, los profesionales de TI con estos conocimientos son difíciles de encontrar debido a la reciente creación de estos sistemas y a sus limitaciones.

Las empresas con repositorios de datos SQL no necesitan científicos de datos avanzados que puedan escribir consultas MapReduce en lenguajes más complejos, como Java. Los profesionales de TI con estos conocimientos en MapReduce están disponibles en un número mucho inferior al de los profesionales con experiencia en SQL. Esto ha llevado a una situación en la que los empleados expertos en estas áreas son difíciles de contratar y costosos de mantener.

Para añadir otra complejidad, tanto las soluciones de datawarehouse tradicionales como los sistemas noSQL emergieron antes de la nube. Las versiones físicas o «migradas a la nube» de estas soluciones nunca reducirán el coste, la complejidad ni los quebraderos de cabeza que

suponen la introducción eficaz de todos sus datos y la extracción de toda la información. Mientras tanto, el acceso a estos sistemas seguirá estando limitado a los pocos miembros del equipo técnico. Estos miembros del equipo solo presionarán para obtener respuestas, no acceso, a muchas más personas de su empresa que trabajan con datos y se basan en ellos para desempeñar su trabajo.



Las implicaciones empresariales de los obstáculos analíticos

Las restricciones tecnológicas que presentan las soluciones noSQL y los datawarehouses heredados no son ningún secreto. El quebradero de cabeza que supone corregir, ajustar, reparar y mantener constantemente un sistema mediocre es un crudo recordatorio de estas limitaciones. Y es probable que todos sus esfuerzos solo ofrezcan mejoras graduales para mantener un sistema que alcanzó su pleno potencial mucho tiempo atrás. Pero ¿qué ocurre con las trabas comerciales que genera esta tecnología?

Cómo la escasez de recursos dificulta el análisis

Los retos provocados por sistemas de datos ineficaces pueden reducirse a tres palabras: escasez de recursos.

Los desafíos que presentan los recursos dificultan su capacidad de sumergirse en la exploración de datos, la inteligencia empresarial, la generación de informes y los análisis predictivos. Sin el tiempo, el dinero ni la inteligencia humana que se necesitan para dedicarse a generar información analítica, es probable que no pueda desarrollar nuevas ideas, productos o procesos a partir de sus datos. En su lugar, se limitará simplemente a ejecutar informes básicos, mientras sus competidores transforman sus negocios con procesos analíticos complejos, tales como el aprendizaje automático, los análisis predictivos y la minería de datos.

La escasez de recursos causada por sistemas de datos ineficaces evita el proceso de innovar productos y prácticas empresariales con el análisis. Simplemente, no hay modo de llevar a cabo experimentos en sus datos y utilizar estos experimentos para generar nuevos datos. En su lugar, se quedará atrapado en el primer paso,

dejando que las ideas se apilen en su cabeza y deseando poder hacer algo con ellas.

La imposibilidad de ampliar el acceso a los datos a todos los usuarios

Las tensiones causadas por el crecimiento exponencial de los datos y las necesidades de una creciente base de usuarios siguen siendo críticas para muchas empresas que se amparan en un datawarehouse como la base de sus análisis.

Debido a los entresijos y a la velocidad glacial del aumento progresivo de la mayoría de las soluciones de datawarehouse, el departamento de Ventas, el departamento de Finanzas y las partes interesadas de los altos cargos directivos, quienes están emocionados por las posibilidades que ofrece el análisis profundo, deberán esperar meses o incluso años para empezar a obtener información de todos sus datos disponibles.

En su lugar, estas partes interesadas seguirán basándose en analistas expertos en TI y datos. Los pocos que posean la capacidad de acceder al datawarehouse ofrecerán respuestas a un pequeño subconjunto de partes interesadas a las que podrán asistir de manera efectiva. Por



supuesto, esto significa que estos pocos analistas dedicarán una gran parte de su día a ejecutar consultas básicas para extraer datos y crear visualizaciones para presentarlos, en lugar de llevar a cabo análisis de datos más avanzados.

Dedicar más tiempo a la infraestructura que a los datos

Muchas organizaciones dedican mucho más tiempo y esfuerzo a mantener la infraestructura de sus datos y sistemas de datos del que invierten en obtener información de los mismos.

Estas necesidades de infraestructura implican el mantenimiento de servidores, la supervisión del rendimiento del sistema y la solución de errores a medida que se producen. A menudo, esto supone tomar decisiones difíciles sobre los datos que son importantes de cargar y las consultas que pueden o no detenerse en aras del rendimiento del sistema. En función del tamaño y de la complejidad de su plataforma para el análisis de datos, algunas organizaciones pueden disponer de varios empleados a tiempo completo, cuya única tarea consiste en controlar y mantener infraestructuras de centros de datos.

Dependiendo de sus protocolos internos y de su arquitectura de datawarehouse actual, también

puede dedicar una gran cantidad de tiempo a extraer, transformar y cargar sus datos en su datawarehouse. Para la mayoría de las organizaciones, ETL puede implicar varios pasos cuando se trata de depurar y modelar los datos para que satisfagan las necesidades del sistema. A esto se añade todo el tiempo dedicado a la infraestructura, la gestión de bases de datos y ETL (cuya complejidad suele ser un resultado de la infraestructura). Esto significa que su equipo de datos desempeñará el papel de policía de tráfico, perdiendo la oportunidad de dedicarse a estrategias y proyectos de análisis significativos que permitan hacer avanzar su empresa y su carrera.



Por qué el datawarehouse sigue siendo relevante

Teniendo en cuenta todos los retos que plantea el datawarehouse, ¿sigue siendo una arquitectura relevante para la era del análisis actual?

La respuesta, sorprendentemente, es sí.

Tal como ocurrió cuando el datawarehouse surgió a principios de los 90, las empresas necesitan un lugar centralizado para albergar y ejecutar consultas en todos sus datos. Además, el aumento de sistemas noSQL, como Hadoop, no ha logrado sustituir al datawarehouse como el modelo dominante.

Los lagos de datos ofrecían un gran potencial dada su capacidad para cargar con facilidad prácticamente cualquier tipo de datos. No obstante, a diferencia de la elegancia organizativa de un datawarehouse, ejecutar una consulta en un lago de datos desorganizado y desgobernado se asemeja en cierto modo a intentar encontrar una talla, un color y una marca específicos de una camisa en una gigantesca tienda de descuentos entre miles de contenedores sin marcar.

El datawarehouse moderno emergió hace tres décadas a partir de dos enfoques ampliamente diferentes para permitir análisis de datos eficaces,

y representa la potencia del almacenamiento de datos tradicional, la flexibilidad de las plataformas de macrodatos y la adaptabilidad de la nube a un coste mucho menor que el de las soluciones anteriores.



La nube es el lugar natural para el datawarehouse moderno

Las ventajas del almacenamiento de datos en la nube moderno sigue eclipsando las preocupaciones iniciales de mover los datos de una organización fuera de su centro de datos local.

Esta migración ha resuelto sin lugar a dudas muchos de los principales problemas de gestionabilidad planteados por los sistemas de datawarehouse locales. Ahora, su organización puede confiar en proveedores basados en la nube para hacer frente a estas necesidades a través de servicios gestionados. Pero no todos los datawarehouses basados en la nube ofrecen la misma arquitectura y tecnología para beneficiarse por completo de las ventajas que presenta la nube.

De hecho, muchos datawarehouses basados en la nube siguen planteando algunos de los mismos obstáculos que representan los sistemas locales para permitir análisis innovadores. Se trata simplemente de soluciones «migradas a la nube» o, incluso, versiones alojadas de datawarehouses físicos. Estos sistemas siguen necesitando a grandes equipos para hacer frente a la gestión de bases de datos y procesos ETL.

Por ejemplo, estos sistemas basados en la nube ofrecen la misma arquitectura tradicional de disponer de datos y procesamiento en el mismo nodo o clúster. Esto significa que el aumento progresivo para atender las crecientes demandas requiere un proceso engorroso para redistribuir los datos antes de que puedan usarse los nuevos recursos. La opción más popular es no hacer nada y seguir actuando como policía de tráfico, proporcionando un escaso desempeño para cada usuario y trabajo del datawarehouse. Mientras tanto, las empresas seguirán ancladas a la espera de hacer realidad su visión de un futuro impulsado por el análisis.

Aunque la nube ha abordado algunos de los retos planteados por los sistemas locales, para beneficiarse realmente de la revolución analítica, las empresas necesitan un datawarehouse diseñado específicamente para la nube que pueda superar los desafíos descritos anteriormente.



El almacenamiento de datos (o datawarehousing) creado para la nube como servicio

Con las características, las funciones y la arquitectura adecuadas, el almacenamiento de datos creado para la nube ofrece un gran potencial para las empresas que buscan superar los retos y los obstáculos a los que se enfrentan a causa de otros sistemas de código abierto y datawarehouse tradicionales.

Solo el datawarehouse moderno creado específicamente para la nube puede ofrecer estas cualidades determinantes para revolucionar el análisis de datos:

- Mejoras exponenciales de rendimiento.
- Ausencia de gestión.
- Escalabilidad infinita (ampliación y reducción)
- Usuarios simultáneos ilimitados con impacto en el rendimiento
- Petabytes de datos estructurados y semiestructurados procesados de forma nativa.

El almacenamiento de datos creado para la nube como servicio ofrece el motor que las empresas necesitan para llevar a cabo análisis exhaustivos de todos sus datos con el fin de desarrollar su negocio. Las ventajas del datawarehouse moderno

se engloban en unas pocas categorías principales.

Adaptabilidad y escalabilidad

Para ofrecer una escalabilidad y una adaptabilidad ideales sin sacrificar el rendimiento, el datawarehouse creado para la nube como servicio combina los mejores aspectos de la arquitectura de disco compartido («shared-disk») y de la arquitectura de nada compartido («shared-nothing»).

Las arquitecturas de disco compartido simplifican la gestión de datos centralizándolos, pero con la desventaja del entorpecimiento del rendimiento entre el almacenamiento y el procesamiento. Por otra parte, las arquitecturas de nada compartido evitan el cuello de botella entre el procesamiento y el almacenamiento, pero con la desventaja

de una gestión de datos complicada, ya que el cambio de tamaño del sistema requiere la redistribución y la reduplicación de los datos. El datawarehouse como servicio ideal impulsa estos dos enfoques con una arquitectura totalmente nueva, que separa los recursos de almacenamiento de los recursos de procesamiento para preservar la velocidad y el rendimiento. Cada consulta que ejecutan los

usuarios, rutina ETL que llevan a cabo o trabajo de desarrollo lanzado utilizaría un clúster de procesamiento independiente frente a una única copia de los datos.

Cada consulta o trabajo se ejecutaría con independencia del resto, y podría ampliarse o reducirse al instante.

En otras palabras, los usuarios obtendrían la velocidad necesaria asegurándose en todo momento de que los datos que ven son precisos y actuales. Sin silos.

Asimismo, puesto que la solución se ofrece como servicio, solo paga por los recursos que usa y cuando los usa. Las empresas que se basan en un datawarehouse físico tradicional o «migrado a la nube» deben planificar con antelación la cantidad máxima de almacenamiento y procesamiento que pueden necesitar en cualquier día del año.

El inconveniente de la arquitectura tradicional es que debe pagar por todos los recursos de almacenamiento y procesamiento que no se utilizan durante los otros 364 días del año. Por otra parte, si las necesidades de su organización superan sus recursos de almacenamiento y procesamiento planificados, el tiempo y el dinero de la ampliación resultarían prohibitivos. Y, si



realiza esta ampliación de la manera tradicional, habrá aumentado de nuevo los recursos que no se utilizan durante el resto del año.

Un verdadero enfoque creado para la nube «como servicio» permite que las empresas amplíen o reduzcan al instante la cantidad de almacenamiento y procesamiento que necesitan del almacén. Esto significa que las empresas siempre cuentan exactamente con los recursos que necesitan sin perder tiempo ni dinero, que podrá emplearse de manera más óptima para desarrollar análisis innovadores.

Variedad de datos

El datawarehouse creado para la nube puede procesar datos estructurados tradicionales y cargar y consultar datos semiestructurados de forma rápida y sencilla.

Un datawarehouse creado para la nube moderno debe identificar automáticamente etiquetas de valor clave y estructuras de conjuntos anidados dentro de los datos semiestructurados (o esquema flexible), y mostrarlos a los analistas a través de SQL. Esto debe ocurrir sin que sea necesario basarse en recursos internos para diseñar y desarrollar previamente estructuras de datos rígidas, y utilizar a continuación herramientas ETL para transformar los datos antes de cargarlos en el almacén. Este enfoque

reduce drásticamente la cantidad de tiempo y los recursos que las empresas necesitarían para transformar y cargar datos semiestructurados, además de mostrar los datos inmediatamente a los analistas a través de SQL, análisis basados en SQL y herramientas BI, una función que no es el fuerte de los sistemas noSQL.

Velocidad

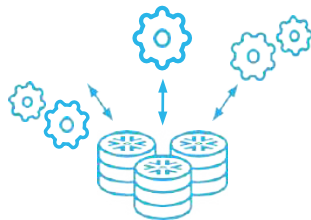
El datawarehouse creado para la nube está optimizado automáticamente para la velocidad. Para una velocidad de consulta óptima, un datawarehouse creado para la nube ofrece a los usuarios la capacidad de iniciar un clúster con procesamiento masivo paralelo (MPP) y adaptar su tamaño para cualquier actividad de desarrollo, trabajo ETL o consulta cuyo tamaño se haya ajustado. Cuando finalice el trabajo o la consulta, el datawarehouse moderno detendrá o desactivará los recursos de procesamiento para que su empresa no deje que el contador siga corriendo en recursos no utilizados. Esto también significa que el datawarehouse creado para la nube moderno puede permitir un número ilimitado de consultas simultáneas, sin que ello afecte al rendimiento.

Para cada consulta o carga individual, un datawarehouse moderno también debe minimizar o eliminar la necesidad de llevar a cabo tareas rutinarias de ajuste del rendimiento, tales como

claves de distribución e indexación, que se requieren con las tecnologías heredadas. También debe incluir un optimizador de consultas inteligente que recopile automáticamente estadísticas sobre las características de los datos como, por ejemplo, las cargas de datos en tiempo real, y que almacene dicha información en un motor de metadatos. A continuación, debe utilizar optimizaciones avanzadas para determinar dinámicamente la ruta más eficaz para acceder a los datos. Este proceso debe tener en cuenta las estructuras de datos y consultas, así como los recursos MPP disponibles en el momento de ejecutar la consulta, lo que acaba con el tiempo normalmente requerido para el ajuste y el análisis manuales, y permite ahorrar una enorme cantidad de espacio (por ejemplo, para índices, etc.) y, por tanto, dinero.



ARQUITECTURA CREADA PARA LA NUBE



DATOS COMPARTIDOS MULTICLÚSTER
 ALMACENAMIENTO CENTRALIZADO ESCALABLE
 CLÚSTER DE PROCESAMIENTO MÚLTIPLE INDEPENDIENTE

ARQUITECTURAS TRADICIONALES



DISCO COMPARTIDO
 ALMACENAMIENTO COMPARTIDO
 CLÚSTER ÚNICO



NADA COMPARTIDO
 ALMACENAMIENTO DESCENTRALIZADO
 LOCAL CLÚSTER ÚNICO

Gestionabilidad

El almacenamiento de datos como servicio acaba con la necesidad de implementar una gran cantidad de recursos humanos y financieros, lo que permite liberar tiempo y talento para centrarse en la generación de información a partir de los análisis de datos y de la inteligencia empresarial. El datawarehouse creado para la nube como servicio acaba con la necesidad de comprar, ejecutar y mantener los muchos elementos de un centro de datos que requieren las soluciones locales tradicionales.

No obstante, la mayoría de los proveedores de datawarehouses en la nube solo gestionan algunos de los aspectos relacionados con la instalación, la gestión y el ajuste de sus soluciones. Estas alternativas suelen adoptar la forma IaaS (infraestructura como servicio) o PaaS (plataforma como servicio).

Solo un verdadero datawarehouse creado para la nube SaaS (software como servicio) ofrece todo el hardware y software como parte de su servicio, incluidos todos los aspectos de la gestión de estos recursos. En el servicio, suelen incluirse actualizaciones de software y hardware, seguridad, disponibilidad, protección de datos y optimización del rendimiento.

Todo esto se traduce en más personal de TI y de análisis de datos con más tiempo para buscar, explorar y experimentar realmente con los datos. Y, gracias a este tiempo adicional, obtendrá los tipos de conocimientos y las ideas necesarios para transformar sus productos, prácticas empresariales y operaciones.

Seguridad

La seguridad sigue siendo la principal preocupación de las organizaciones que migran datos a la nube. A continuación, le indicamos algunas de las medidas estándar de la industria que debe ofrecer un datawarehouse en la nube:

1. Cifrar datos en tránsito y en reposo

Si un usuario no autorizado logra acceder a sus datos, no debe poder leerlos. Punto y pelota. El datawarehouse en la nube moderno debe proteger los datos en tránsito y en reposo, siempre que se transmitan a través de una red o se guarden en un disco. Esto incluye archivos de datos almacenados persistentemente, resultados de consultas y el contenido de una caché de disco local. Asimismo, una solución de datawarehouse en la nube moderna debe utilizar los últimos algoritmos de cifrado estándar de la industria. El Estándar de



Cifrado Avanzado (AES), con claves de 128 bits, es la práctica recomendada mínima para el cifrado simétrico. Para una seguridad mejorada, los datawarehouses en la nube más robustos utilizan AES-256.

2. Gestión de claves

Cualquier datawarehouse, debe limitar la cantidad de datos cubierta por una clave de cifrado individual y el tiempo que se debe utilizar la clave. Se trata de una práctica recomendada de la industria que se materializa mediante la rotación de claves y la regeneración

de claves de datos.

La rotación de claves es un método que permite generar periódicamente una nueva clave de cifrado para proteger datos recién insertados.

La regeneración de claves es la capacidad de volver a los datos almacenados previamente, cifrarlos de nuevo con nuevas claves de cifrado recién generadas y deshacerse a continuación de las claves de cifrado anteriores.

3. Autenticación Multifactor

Cómo práctica recomendada, cualquier persona que acceda a los datos de un datawarehouse en la nube debe llevar a cabo esta acción utilizando la autenticación multifactor (MFA). Tras iniciar sesión con un nombre de usuario y una contraseña, el usuario necesitará un segundo mecanismo de autenticación, que puede tratarse de un código aleatorio generado por una aplicación en el smartphone de un usuario.

4. Verificación de terceros

Un proveedor de datawarehouses en la nube debe cumplir los siguientes estándares, en función del tipo de datos que pretenda almacenar:

SOC 2: la finalidad de un informe SOC 2 es evaluar los sistemas de información de una organización en lo que respecta a la seguridad, la disponibilidad, la integridad del procesamiento, la confidencialidad o la privacidad.

HIPAA: la Información Sanitaria Protegida (PHI) está sujeta a las normas de seguridad y privacidad en virtud de la Ley de Portabilidad y Responsabilidad de los Seguros Médicos (HIPAA) de Estados Unidos.

PCI: el cumplimiento de la industria de tarjetas de pago (PCI) consiste en el respeto de un conjunto de estándares de seguridad específicos.

desarrollados para proteger información de tarjetas de crédito durante y después de una transacción financiera. Todas las marcas de tarjetas requieren el cumplimiento PCI.

Las ventajas empresariales de implementar un datawarehouse en la nube

Hasta dónde puede llegar su empresa con el análisis de datos está determinado en gran medida por el motor que impulsa dichos análisis.

A medida que la competencia encuentra formas más efectivas de almacenar datos y aumentar sus análisis, las organizaciones rezagadas se quedarán atrás, no solo en términos de innovación, sino también en términos de ingresos y rentabilidad. La norma general, pero desafortunada, es que los analistas de datos dedican el 80 % de su tiempo a recuperar y organizar datos, y el 20 % restante a llevar a cabo realmente análisis de datos. Si dicha proporción se invirtiera, ¿hasta qué punto podría decirse que su organización trabaja con datos?

Si está anclado en la ejecución de consultas lentas para generar informes básicos, ¿cómo podrá beneficiarse alguna vez de la minería de datos, los análisis predictivos, el aprendizaje automático o lo que sea que depare el futuro del análisis? Después de todo, el conocimiento que adquiere de un enfoque desinhibido para analizar sus datos puede representar la inteligencia que necesita para transformar su negocio.

«Los conocimientos profundos suelen residir en las fronteras», afirma McKinsey. «Al igual que la consideración de datos “blandos” puede desvelar nuevos conocimientos, la combinación de las fuentes de información de una determinada persona puede definir aún más dichas revelaciones».

En los próximos años, las empresas que prosperen serán las que se adapten rápidamente a los cambios que las rodean. Para ver dichos cambios con claridad, necesita una plataforma sólida que impulse sus análisis de datos. Por tanto, debe convertirse en un experto capaz de ayudar a convertir su organización en una empresa basada en datos.

Las empresas que toman decisiones basadas en datos satisfarán a los clientes, agilizarán sus operaciones y liderarán sus industrias

El datawarehouse creado para la nube como servicio ofrece esta base. Al maximizar la velocidad, minimizar los costes y liberar de carga de trabajo a los recursos humanos, las empresas podrán obtener el tiempo y la flexibilidad que necesitan para beneficiarse del valor extraído de sus datos.

Caso de éxito: De la implementación a la obtención de beneficios en menos de un año

Con la existencia de un datawarehouse adecuado como base, un futuro impulsado por análisis potentes es fácilmente imaginable. Pero puede que una sencilla historia de velocidad en la generación de valor y retorno de la inversión (ROI) ilustre mejor el valor transformador del datawarehouse creado para la nube como servicio.

PDX, un proveedor de análisis SaaS para miles de farmacias, experimentó muchas de las frustraciones de su datawarehouse heredado. Tras una demostración conceptual, optó por implementar un datawarehouse creado para la nube con toda la velocidad, escalabilidad y gestionabilidad descritas anteriormente.

Gracias a una rápida implementación, la empresa combinó su nuevo datawarehouse con una herramienta analítica y empezó a desarrollar un servicio de análisis de datos para sus clientes farmacéuticos. Este servicio ofrecía a las farmacias información sobre sus clientes que no era posible con el datawarehouse anterior de PDX. En un plazo de ocho meses, este servicio había crecido con tal éxito que la solución SaaS de PDX empezó a generar beneficios. En otras palabras, en menos tiempo del que se tardó en implementar un sistema de datawarehouse heredado, esta empresa ya había puesto en funcionamiento su datawarehouse creado para la nube y estaba utilizando un conglomerado de fuentes de datos de una única ubicación al mismo tiempo que obtenía beneficios.





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

