

Snowflake:
Una plataforma
de datos en la
nube para todas
sus necesidades
de análisis



Contenido

Introducción	03	Lo que permite la plataforma de datos en la nube de Snowflake	13
Cómo protege Snowflake sus datos en la nube	08	Conclusión	14
Intercambio de datos de Snowflake	10	Fuente de Snowflake	15
Arquitectura de datos moderna con Snowflake	12	Acerca de PowerData	16

Introducción

Para 2023, IDC predice que el 50% de los ingresos de los proveedores de software para datos y análisis provendrán de las nubes públicas, y que las implementaciones en la nube Gartner va más allá y predice que el 75% de todas las bases de datos se desplegarán o migrarán a una plataforma en la nube en 2022.

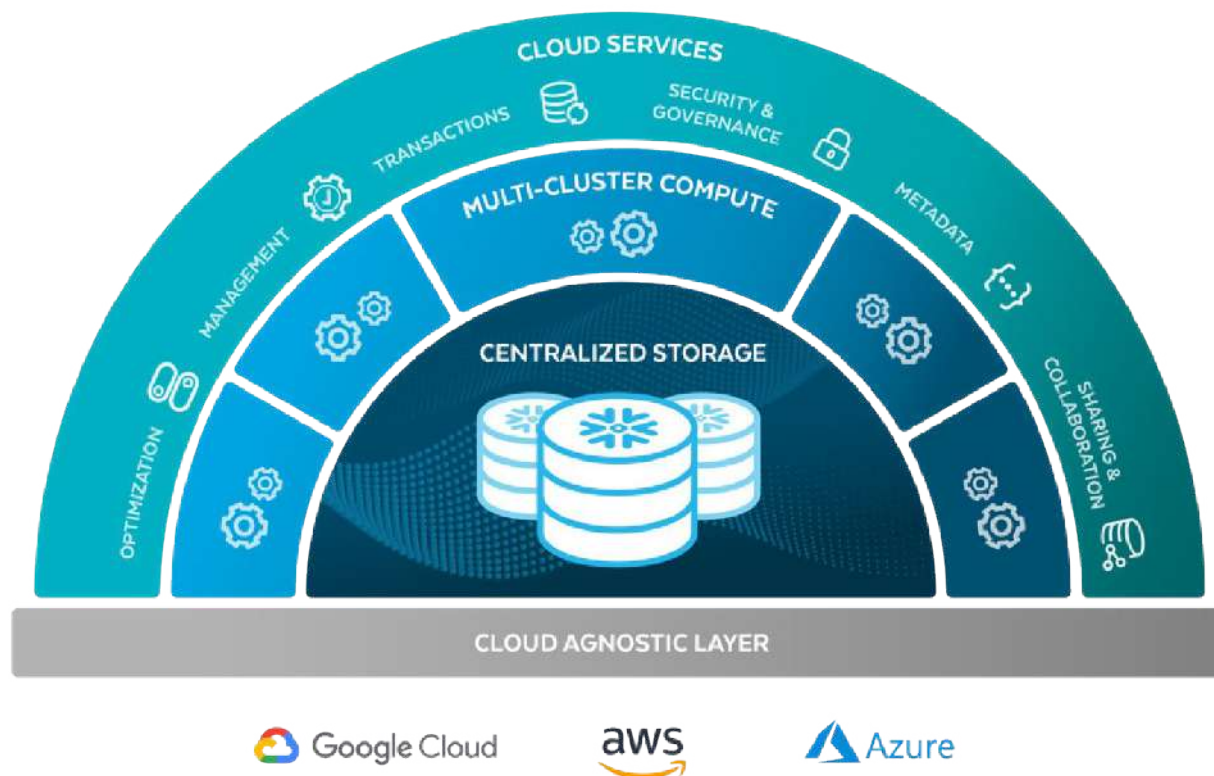
Sin embargo, las organizaciones con visión de futuro quieren una plataforma de datos en la nube que permita una estrategia a largo plazo para maximizar sus activos de datos, basada en una plataforma extensible, multirregional y multi-nube que potencie todo tipo de cargas de trabajo de datos. No quieren trasladar los datos y las cargas de trabajo analíticas a la nube de forma aleatoria. En respuesta, Snowflake ha creado una plataforma de datos en la nube que permite lo mejor del almacenamiento de datos moderno, lo mejor de los lagos de datos, las capacidades de ciencia de datos más avanzadas y mucho más.

La plataforma de datos en la nube de Snowflake le permite consolidar los análisis de los almacenes de datos, los mercados de datos y los lagos de datos en una única fuente de verdad que impulsa múltiples tipos de aplicaciones empresariales. También agiliza la forma en que diversos equipos comparten los datos gobernados, interna y externamente, para colaborar en todos los datos sin tener que copiarlos y trasladarlos de un lugar a otro. Esto simplifica el intercambio de datos, permite establecer las restricciones de acceso a los datos pertinentes y minimiza los problemas de gobierno de datos y de cumplimiento de la normativa que supone la gestión de varias copias de los datos y el cumplimiento de la normativa de la gestión de múltiples copias de los mismos conjuntos de datos.



Snowflake también le permite almacenar sus datos en cualquier región, en cualquier nube, incluidas las ofertas populares de Amazon, Microsoft y Google (véase la Figura 1).

La plataforma automatiza todo, desde el almacenamiento y el procesamiento de datos hasta la gestión de transacciones, la seguridad, la gobernanza y la gestión de metadatos



Arquitectura Informática Multiclúster y Multicloud

Mientras que las arquitecturas de datos tradicionales son difíciles de escalar, una de las principales diferencias de Snowflake le permite escalar la cantidad exacta de recursos informáticos hacia arriba y hacia abajo, de forma instantánea e independiente. Todo el procesamiento de datos se lleva a cabo con la cantidad de potencia de cálculo que usted elija para cada carga de trabajo, que puede reducir con la misma rapidez. La arquitectura única de Snowflake le permite procesar enormes cantidades de datos con la máxima velocidad y eficiencia al dedicar una potencia de cálculo independiente a cada carga de trabajo, por lo que no hay impacto en el rendimiento. Puede especificar el número de clusters de computación que desea utilizar o dejar que el servicio escale automáticamente. Diseñado para la eficiencia, Snowflake determina con precisión lo que se requiere de la capa de almacenamiento, eliminando escaneos innecesarios para satisfacer las consultas de la manera más eficaz posible. Los datos se almacenan en caché junto con los resultados de la consulta para mejorar el rendimiento de futuras consultas. Esta arquitectura permite ejecutar múltiples cargas de trabajo en varios equipos sin contención de recursos, maximizando el rendimiento y la eficiencia.

Flexible, escalable y rentable

La plataforma de datos en la nube de Snowflake se adapta a miles de usuarios, facilitando la colaboración en toda la empresa. Es una solución ideal para potenciar los análisis de autoservicio, incluidos los informes operativos, las consultas ad hoc y la toma de decisiones en tiempo real. El aislamiento de recursos garantiza que todos los usuarios tengan una experiencia consistente, mientras que la facturación por segundos de Snowflake significa que cada usuario y grupo de trabajo paga sólo por los recursos precisos utilizados. Es fácil ampliar y escalar cada aplicación analítica, pero pagando sólo por la potencia de cálculo que utilice y la capacidad de almacenamiento que necesite. La arquitectura avanzada de Snowflake, que se ofrece como un servicio y con una funcionalidad consistente en múltiples regiones y nubes, permite una elasticidad instantánea y una plataforma cohesiva para servir a todos los tipos de usuarios y cargas de trabajo de forma consistente. manera consistente. Entre bastidores, la plataforma automatiza de forma transparente el aprovisionamiento de servidores, la gestión de datos, la protección de datos, la seguridad de datos y el ajuste del rendimiento. Esto libera a su equipo para que se centre en obtener el máximo valor de sus datos.

Buenas Prácticas

La arquitectura única de Snowflake separa los recursos de almacenamiento de los recursos informáticos: todas las cargas de trabajo pueden aprovechar simultáneamente sólo la potencia informática que necesitan, cuando la necesitan, sin contención esto.

Servicios de plataforma de datos compartidos Multiclúster

Los servicios de datos compartidos multiclúster de Snowflake incluyen un conjunto integrado de servicios para autenticar las sesiones de los usuarios, gestionar los recursos, reforzar la seguridad, compilar las consultas y garantizar la integridad de todas las transacciones. Esta capa de servicios está formada por recursos informáticos sin estado que se ejecutan en múltiples zonas de disponibilidad. Utiliza un almacén de metadatos distribuido y de alta disponibilidad para la gestión global del estado. ¿Suenan complicado? No lo es. Con la plataforma de datos en la nube de Snowflake, altamente eficiente y sin problemas, no hay que comprar hardware, ni software que mantener, ni infraestructura que integrar. A medida que sus cargas de trabajo se ejecutan, la plataforma de datos determina la forma más eficiente, rentable y de alto rendimiento para procesar sus datos. Gracias a esta moderna arquitectura, la plataforma de datos en la nube Snowflake puede manejar muchas cargas de trabajo simultáneamente, incluso cuando residen en varias nubes y regiones, con un rendimiento excepcional.

Incorporación continua de datos

La plataforma de datos en la nube Snowflake incluye un servicio de ingestión sin servidor llamado Snowpipe que carga los datos de forma asíncrona en su entorno de almacenamiento en la nube. Los conectores y adaptadores estándar le permiten ingerir fácilmente flujos de eventos de Kafka y otros sistemas de mensajería, mientras que los flujos y las tareas de Snowflake facilitan la programación de cargas de datos para trabajos SQL. Puede ingerir datos en el repositorio y compartirlos con una base de clientes global, sin tener que configurar conductos de extracción, transformación y carga (ETL) o intercambiar datos entre regiones. La plataforma transforma automáticamente los datos en el tipo y la forma necesarios para cada tabla de destino. Un conector Apache Kafka le permite transmitir continuamente registros JSON para su almacenamiento y análisis.

Buenas Prácticas

La capa de servicios de la plataforma de datos en la nube Snowflake unifica la seguridad, la gobernanza y la gestión de metadatos, protegiendo sus datos, optimizando el rendimiento de las cargas de trabajo, eliminando la contención de recursos y garantizando la coherencia de las transacciones.

Gestión robusta de transacciones y partición de datos

Garantizar la integridad de las transacciones de las bases de datos es esencial en muchos sectores. Esto es especialmente importante para las organizaciones que manejan datos sensibles al tiempo, como las empresas de servicios financieros que realizan transacciones monetarias y las empresas de fabricación que ejecutan procesos de producción en tiempo real. Snowflake aplica las propiedades de atomicidad, consistencia, aislamiento y durabilidad (ACID) de las bases de datos para garantizar la consistencia de las transacciones incluso en caso de errores imprevistos, fallos de alimentación y otros percances. El cumplimiento de las propiedades ACID garantiza la precisión de la base de datos y la integridad de las transacciones, mientras que el microparticionamiento mejora el rendimiento de la base de datos. A medida que se cargan los datos de datos se cargan, Snowflake los divide de forma transparente en microparticiones, unidades contiguas de almacenamiento organizadas en forma de columnas. Esta estructura permite una poda extremadamente granular de tablas muy grandes, que pueden estar compuestas por millones de microparticiones para mejorar significativamente el rendimiento de las consultas.

Snowflake almacena las particiones en una capa de almacenamiento central a la que pueden acceder todos los nodos de cálculo. Durante una consulta, selecciona automáticamente el método de distribución óptimo para las particiones en función del tamaño actual del conjunto de datos.

Seguridad y cumplimiento de la normativa líderes en el sector

La plataforma de datos en la nube Snowflake está construida sobre una base de seguridad multicapa que incluye encriptación, control de acceso, supervisión de la red y medidas de seguridad física, junto con una completa supervisión, alertas y prácticas de ciberseguridad. Todos los aspectos de la plataforma están orientados a proteger sus datos, tanto en tránsito como en reposo. Tanto si sus datos residen en una o en muchas ubicaciones, tiene que proteger y controlar el flujo de datos en cada punto. Snowflake se toma esto muy en serio: las tecnologías de seguridad se entretejen en todas las facetas de la plataforma (véase la Figura 2 en la página siguiente). Por ejemplo, el control de acceso basado en las funciones garantiza que los usuarios sólo puedan acceder a los datos que están autorizados a ver.

Buenas Prácticas

La plataforma de datos en la nube de Snowflake admite la carga y el análisis de datos en formatos de datos mixtos con total integridad transaccional, garantizando la precisión de todas las transacciones de la base de datos y asegurando un óptimo para todo tipo de consultas.

Estos mecanismos de control pueden aplicarse a todos los objetos de la base de datos, incluidas las tablas, los esquemas y cualquier extensión virtual de un conjunto de datos. Los procedimientos de autenticación multifactor pueden emitir una verificación secundaria, como un código de

seguridad de un solo uso enviado al teléfono móvil del usuario.

Sin embargo, toda esta seguridad integral no se interpone en la usabilidad. Los procedimientos de inicio de sesión único y las tecnologías de

autenticación federada facilitan el acceso a un almacén de datos, un lago de datos u otro servicio analítico directamente desde esas aplicaciones.

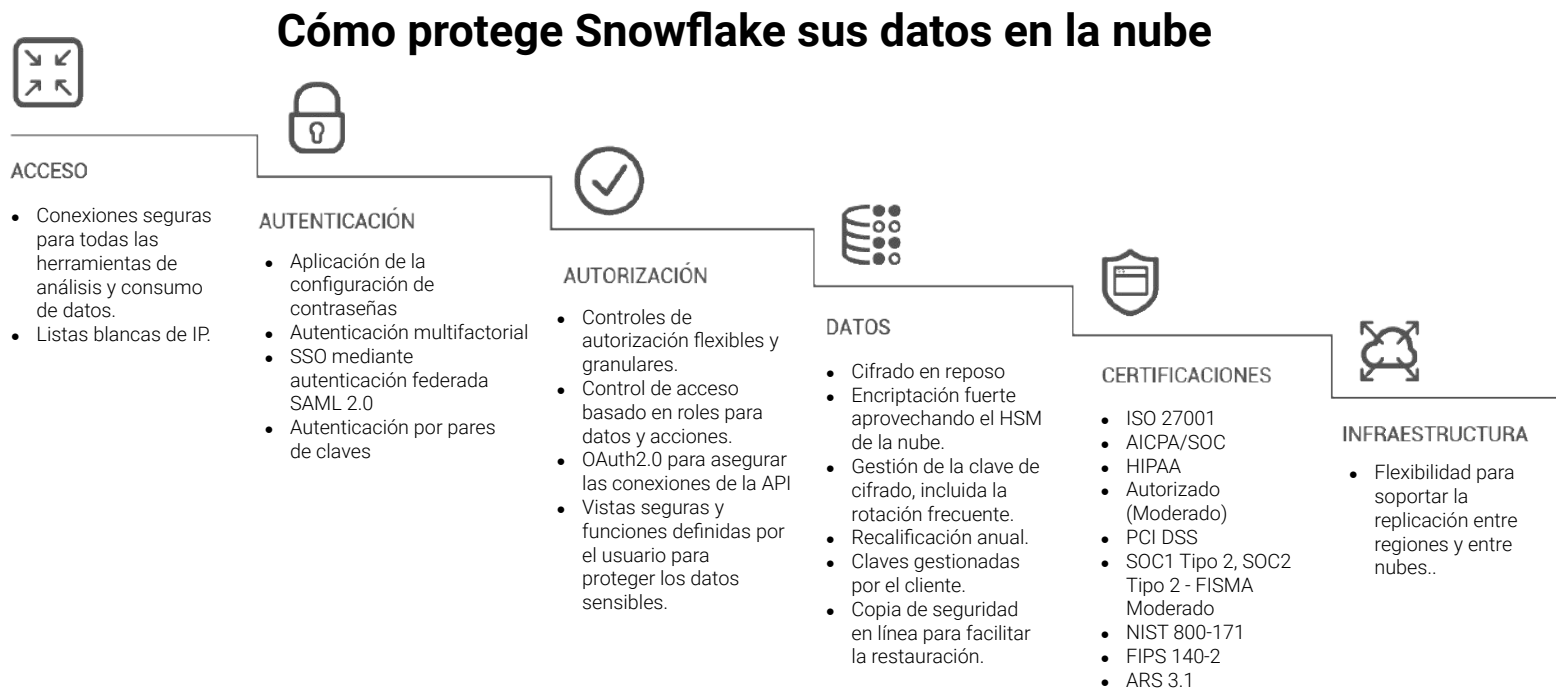


Figura 2: Las tecnologías de seguridad se entrelazan en todas las facetas de la plataforma.

Gestión de metadatos para reforzar la precisión y la coherencia

Snowflake reduce la complejidad al consolidar datos en un único repositorio que puede acomodar datos estructurados o semiestructurados de cualquier fuente generados en las instalaciones o en la nube-, incluyendo bases de datos OLTP, aplicaciones empresariales, aplicaciones de terceros, datos de registro web y datos IoT. Un sólido servicio de metadatos abarca todo el sistema. Las consultas se compilan dentro de la capa de servicios, y los metadatos se utilizan para determinar las columnas de la micropartición que deben analizarse. Esto permite rastrear la procedencia de los datos, quién los ha tocado y cómo se relacionan los distintos conjuntos de datos. Las tablas externas almacenan metadatos a nivel de archivo, como las rutas de los archivos, los identificadores de versión y la información de partición. Esto permite a los usuarios de la empresa consultar los datos a través de múltiples cargas de trabajo de forma coherente, como si todo surgiera de una única base de datos.

Almacenamiento elástico y multifacético

Ya sea que esté importando datos relacionales de aplicaciones empresariales, datos de registro de actividades de sitios web, datos de IoT de sensores de equipos o datos de medios sociales de redes de consumidores, el volumen y la complejidad de las fuentes de datos actuales pueden abrumar rápidamente a una plataforma de datos convencional. Necesita una capa de almacenamiento elástica para almacenar fácil y eficazmente almacenar todos sus datos y hacerlos disponibles al instante, independientemente de la escala.

La plataforma de datos en la nube de Snowflake maneja tanto datos estructurados como semiestructurados y puede abarcar múltiples regiones y nubes, consolidando todos sus datos en una única fuente de verdad. Por ejemplo, además de los datos relacionales de las aplicaciones empresariales, puede utilizar este repositorio para almacenar datos semiestructurados como JSON, Avro y XML, junto con tipos de datos de código abierto como Apache Parquet y ORC, y todo ello a escala de petabytes. Construida sobre un versátil almacenamiento blob, esta capa de almacenamiento contiene sus datos, tablas y resultados de consultas.

Buenas Prácticas

La plataforma de datos en la nube de Snowflake protege sus datos tanto en tránsito como en reposo, haciendo hincapié en el cifrado, el control de acceso, el almacenamiento de datos y la infraestructura física junto con prácticas de ciberseguridad exhaustivas.

Compartir e intercambiar datos de forma segura

Los métodos tradicionales de intercambio de datos requieren que cualquier organización que comparta datos los copie y los envíe a sus consumidores de datos. Estos métodos engorrosos, costosos y arriesgados se basan en el intercambio de datos estáticos, que rápidamente se vuelven obsoletos y deben ser actualizarse continuamente con las versiones más recientes. Enfoques anticuados

como estos suponen costes innecesarios y requieren un esfuerzo adicional, pero carecen de la capacidad de asegurar la información sensible o de prevenir las violaciones de datos.

La plataforma de datos en la nube de Snowflake cuenta con Snowflake Data Sharing, que permite a los miembros autorizados de un ecosistema en la nube acceder a versiones en vivo y de sólo lectura de los datos. Esto le permite compartir fácilmente y de forma segura subconjuntos de sus datos, así

como recibir datos compartidos de forma segura y gobernada. Por ejemplo, Snowflake Data Sharing permite a los proveedores de datos compartir instantáneamente datos con proveedores, socios de la cadena de suministro, socios logísticos, clientes y muchos otros componentes. También puede establecer intercambios de datos y otros servicios de intercambio de datos con Snowflake para monetizar sus datos, convirtiendo su almacén o lago de datos en un centro de beneficios (véase la figura 3).

Intercambio de datos de Snowflake

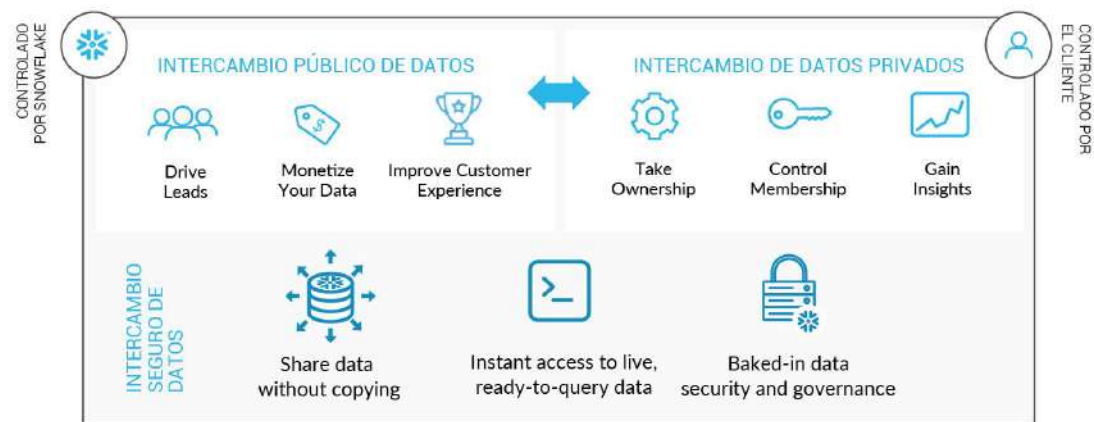


Figura 3: Snowflake Data Sharing le permite compartir datos externamente a través del Public Data Exchange, y crear su propio Private Data Exchange con clientes, proveedores y otros socios comerciales.

Alto rendimiento y carga de trabajo múltiple

Las tablas externas de Snowflake funcionan con datos almacenados directamente en su lago de datos de Amazon S3, Azure Blob Storage o Google Cloud Platform. Las vistas materializadas en estas tablas externas le permiten materializar todo o sólo la parte del conjunto de datos que utiliza con más frecuencia, eliminando la necesidad de construir una capa ETL o un pipeline de orquestación. En lugar de transferir físicamente los datos a los consumidores internos o externos, la tecnología de compartición de datos de Snowflake permite el acceso de sólo lectura a una parte gobernada del conjunto de datos en vivo a través de SQL. Los datos compartidos Los datos compartidos están disponibles inmediatamente para ser consultados sin necesidad de transformación. Los datos nunca salen de la cuenta del proveedor de datos. Los consumidores de datos no incurrir en costes de almacenamiento y no necesitan preocuparse por los procesos ETL para hacer uso de los datos. Los flujos y las tareas de Snowflake facilitan la programación de las cargas de datos para los trabajos de SQL. La plataforma transforma automáticamente los datos en el tipo y la forma necesarios para cada tabla de destino. Un conector Apache Kafka le permite transmitir continuamente registros JSON para su almacenamiento y análisis.

Replicación global de datos entre múltiples nubes

Tener una huella global ayuda a garantizar la recuperación de desastres y a mejorar las prácticas de continuidad del negocio. También amplía sus opciones para compartir datos con una base de clientes global. La plataforma de datos de Snowflake facilita la compartición de datos entre regiones geográficas y entre múltiples nubes y regiones, incluidas las ofrecidas por Amazon Web Services, Microsoft Azure y Google Cloud Platform. Estas capacidades de replicación global facilitan que sus unidades de negocio, y sus socios, colaboren entre sí a través de un ecosistema global para el análisis y el intercambio de datos. La avanzada tecnología de replicación de Snowflake le permite compartir fácilmente los datos entre regiones y nubes sin tener que configurar conductos de datos, copiar datos o resolver diferencias en los protocolos de seguridad entre las distintas nubes. La replicación de datos en varias regiones y nubes no sólo simplifica el intercambio de datos, sino que facilita la expansión global y refuerza la continuidad del negocio. La plataforma replica automáticamente las bases de datos y las mantiene sincronizadas en todas las regiones y nubes, impulsando la disponibilidad, automatizando la conmutación por error y garantizando el acceso y la recuperación instantánea de bases de datos de cualquier tamaño.

Buenas Prácticas

Con la capacidad de compartir datos de forma segura de Snowflake, no tiene que copiar o mover sus datos. Los datos están vivos, listos para usar y siempre frescos.

Extensibilidad para el futuro

La plataforma de datos en la nube Snowflake es extensible, lo que significa que es accesible a otras soluciones autorizadas. La plataforma proporciona una base sólida para muchos tipos de datos y aplicaciones. Como repositorio extensible y servicio de datos unificado, la plataforma puede trabajar con los datos allí donde residen, eliminando la necesidad de mover los datos desde un almacén de datos o un lago de datos. Esto

permite gestionar datos en muchos lugares dispares y de muchas cargas de trabajo diversas, pero desde un único servicio, independientemente de la fuente y el tipo de datos que utilice cada solución o de cómo se carguen, integren y transformen. Este versátil servicio en la nube maximiza el valor de sus datos y de todos los servicios que se conectan a ellos, ahora y en el futuro (véase la Figura 4 en la página siguiente).

Arquitectura de datos moderna con Snowflake



Figura 4: Snowflake potencia y amplía su arquitectura de datos, de modo que reunir fácilmente todos sus datos en una única ubicación y obtener toda la información de esos datos.

Buenas Prácticas

La plataforma de datos entre nubes de Snowflake permite el movimiento libre y seguro de los datos en cualquier parte del mundo, a la vez que permite seleccionar los proveedores de almacenamiento en la nube que satisfacen las necesidades de cada aplicación o unidad de negocio.

Una base sólida para cualquier carga de trabajo

La mayoría de las empresas se enfrentan a una enorme diversidad de datos. Los profesionales de TI deben racionalizar los datos que llegan desde una variedad de fuentes, regiones y nubes. Algunos de los datos están estructurados, otros están semiestructurados, y no hay una única fuente de verdad a partir de la cual puedan correlacionar los análisis de forma fiable.

La plataforma de datos en la nube de Snowflake pone orden en este caótico panorama. Da cabida a todos estos tipos de datos, a escala ilimitada, y permite almacenar diferentes tipos de datos en sus formatos nativos, sin necesidad de complejas transformaciones.

Snowflake admite la transmisión en tiempo real, así como la carga por lotes. Esta arquitectura ofrece una base sólida para ejecutar cualquier carga de trabajo, incluidos los almacenes de datos, los lagos de datos y muchos tipos de aplicaciones de ingeniería y ciencia de datos.

Disponer de un repositorio unificado alimentado por una amplia capa de servicios aporta coherencia a todo tipo de proyectos analíticos. Snowflake consolida todos los recursos necesarios en una plataforma cohesiva para permitir muchas cargas de trabajo de datos que las organizaciones necesitan para agilizar su negocio, servir mejor a sus clientes y revelar nuevas oportunidades de mercado. (véase la figura 5 más abajo).

Buenas Prácticas

Snowflake le permite almacenar diversos tipos de datos en sus formas nativas, sin crear nuevos silos de datos. La elasticidad automática y casi infinita libera los recursos que necesita, cuando los necesita.

Lo que permite la plataforma de datos en la nube de Snowflake



Figura 5: Snowflake permite muchos casos de uso y cargas de trabajo para que su organización pueda aprovechar la potencia de una única solución para el análisis de datos

Conclusión

Con una seguridad y gobernanza unificadas, la capacidad de almacenar tipos de datos en su formato nativo y la seguridad transaccional integrada, la estandarización de la plataforma de datos en la nube de Snowflake le permite cambiar su enfoque de la gestión de la infraestructura a la gestión de los datos, y obtener toda la información posible de esos datos. La arquitectura avanzada de Snowflake, que se ofrece como un servicio y con una funcionalidad consistente en múltiples regiones y nubes, permite una elasticidad instantánea y casi infinita y proporciona una plataforma cohesiva para servir a todos los tipos de usuarios y cargas de trabajo de forma consistente. Con una única plataforma para obtener toda la información de todos sus datos, su organización puede tomar decisiones basadas en datos que sirvan mejor a sus clientes y a su negocio, ahora y en el futuro.

Fuentes de Snowflake

Esta guía es una traducción al español de la versión original “[Snowflake: One Cloud Data Platform for All Your Analytics Needs](#)” 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

