

VALIDACIÓN ECONÓMICA

Las ventajas económicas de Google Cloud BigQuery frente a soluciones de EDW alternativas basadas en la nube

Un EDW verdaderamente sin servidor basado en la nube que ofrece simplicidad operativa, escalabilidad y flexibilidad sin igual y ahorro de costos de hasta un 54 % en comparación con EDW alternativos basados en la nube

Por Aviv Kaufmann, director de Práctica y analista principal de validación económica
Enterprise Strategy Group

Febrero del 2024

Índice

| | |
|--|-----------|
| Introducción | 3 |
| Desafíos..... | 3 |
| La solución: Plataforma de datos empresariales sin servidor Google Cloud BigQuery | 4 |
| BigQuery frente a servicios de EDW alternativos basados en la nube..... | 6 |
| Validación económica de Enterprise Strategy Group | 8 |
| Descripción general del valor económico: BigQuery frente a soluciones de EDW alternativas basadas en la nube | 8 |
| Eliminación de la inversión inicial y la planificación del tamaño de implementación | 8 |
| Reducción de los gastos operativos | 12 |
| Mejora de la agilidad comercial y reducción del costo de la administración diaria de EDW | 13 |
| Consideraciones de migración..... | 16 |
| Análisis de Enterprise Strategy Group..... | 16 |
| Gasto inicial en la nube (inversión de capital prepagado)..... | 18 |
| Costos de la nube a pedido..... | 19 |
| Costos administrativos | 19 |
| Problemas que se deben considerar | 20 |
| Conclusión | 21 |



Validación económica: Resumen de hallazgos clave

Beneficios validados de Google Cloud BigQuery



De un 28 % a un 54 % menos de costo total de propiedad (Total Cost of Ownership, TCO) en comparación con las ofertas de EDW en la nube alternativas

- No se requiere inversión inicial
- Precios justos y flexibles
- Verdadero diseño sin servidor con la más alta escalabilidad y simultaneidad
- Almacenamiento de BigQuery y escalamiento automático optimizados para rendimiento y costo
- Compatibilidad con IA/AA y GenAI

- Eliminación de la inversión inicial y la planificación:** Los precios flexibles y predecibles de procesamiento y almacenamiento eliminan la necesidad de pagar meses o incluso años antes para reducir el gasto en la nube y elimina las conjeturas y la planificación necesarias para ajustar el tamaño de los recursos.
- Reducción de los gastos operativos:** Elimina la necesidad de administrar nodos de almacén de datos empresariales (Enterprise Data Warehouse, EDW) virtuales, así como la necesidad de monitorear, solucionar problemas, actualizar, ajustar y planificar el crecimiento.
- Mayor agilidad comercial:** La escala aumenta o disminuye según la necesidad para satisfacer las demandas comerciales cambiantes, lo que permite a las organizaciones actuar rápidamente cuando se presenten nuevas oportunidades sin necesidad de planificar los requisitos de configuración, pausar las bases de datos o iniciar instancias dedicadas de almacenes para cada organización.

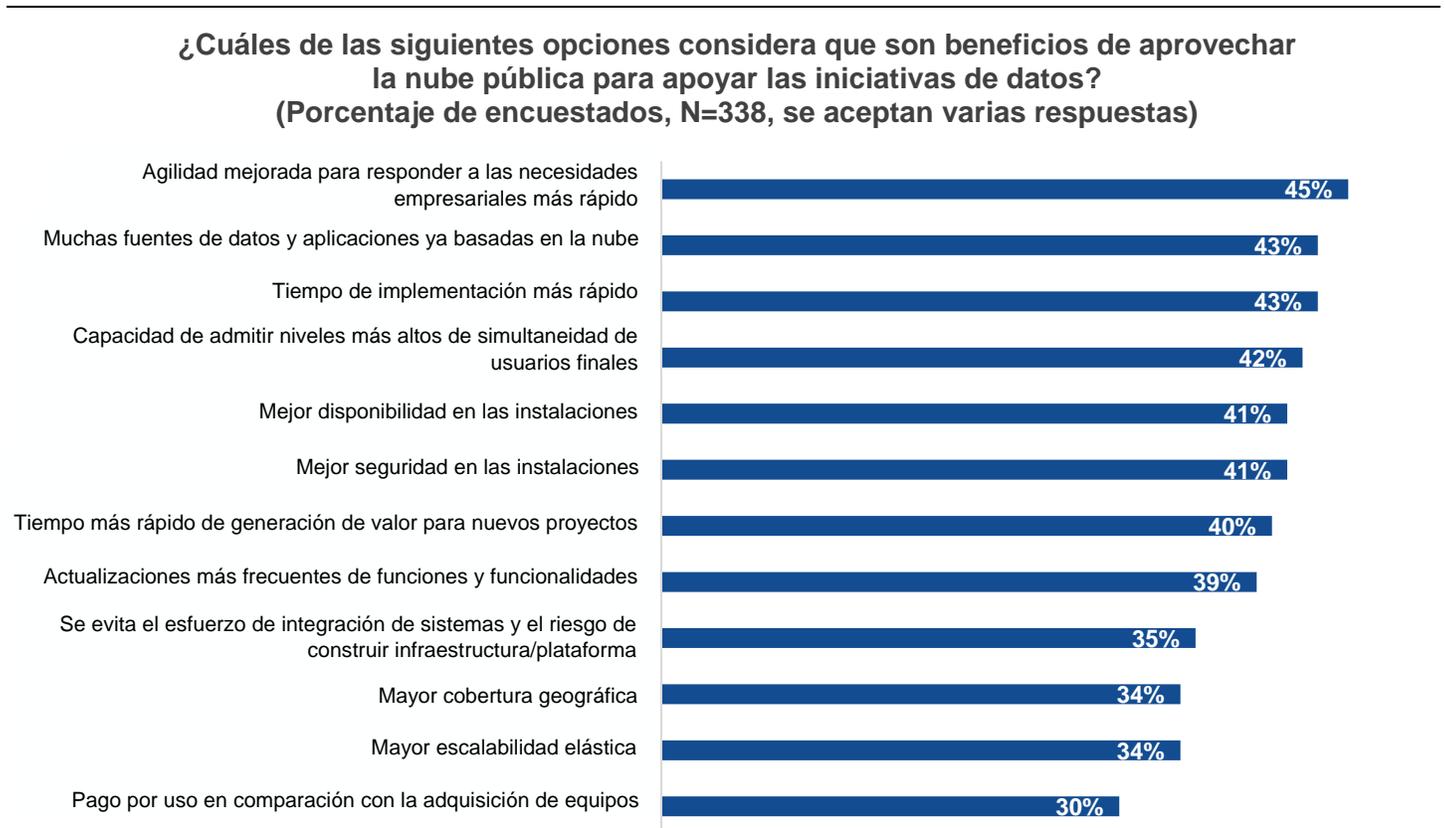
Introducción

Esta validación económica del Enterprise Strategy Group (ESG) de TechTarget se centra en los beneficios cuantitativos y cualitativos que las organizaciones pueden esperar obtener con Google Cloud BigQuery cuando se compara con servicios alternativos de almacén de datos empresariales (EDW) basados en la nube.

Desafíos

Las organizaciones modernas impulsadas por datos recopilan datos de una creciente variedad de fuentes, dispositivos y ubicaciones. La única línea de corte clara entre los datos designados para plataformas de datos estructuradas, no estructuradas, administradas y de código abierto ha sido borrosa, y las organizaciones están adoptando un enfoque más eficiente y consolidando sus estrategias de datos en la nube. El desarrollo de una nube de datos proporciona agilidad de TI, escalabilidad, consolidación simple de fuentes de datos y acceso a una incomparable variedad de herramientas y servicios de IA y aprendizaje automático (AA) diseñados para extraer información e inteligencia. Un estudio de investigación de ESG descubrió que aprovechar la nube pública para respaldar las iniciativas de datos da como resultado una agilidad mejorada para responder a las necesidades de la empresa más rápido, la proximidad a las fuentes de datos y aplicaciones que ya se encuentran en la nube, tiempos de implementación más rápidos, una capacidad para admitir niveles más altos de simultaneidad de usuarios finales, y mayor disponibilidad y seguridad (consulte la Figura 1).¹

Figura 1. Los beneficios principales de aprovechar la nube pública para apoyar las iniciativas de datos



Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

¹ Fuente: Enterprise Strategy Group Report, [Cloud Analytics Trends](#), marzo de 2022. Todas las referencias de investigación y gráficos de ESG en esta validación económica se tomaron de este informe de investigación a menos que se indique lo contrario.

Nuestra investigación también demuestra que, si bien el 86 % de las organizaciones cree que son eficientes para actuar en base a nueva información de datos, solo el 4 % de estas organizaciones puede extraer información de sus datos en tiempo real y un sorprendente 45 % de las organizaciones afirma que les toma semanas o más tiempo obtener esta información. El almacén de datos en las instalaciones tradicional ya no es lo suficientemente ágil o escalable para mantenerse al día con las demandas cambiantes de las organizaciones modernas impulsadas por datos. Una estrategia que incorpora almacenes de datos en la nube, lagos de datos, IA, AA y transmisión y procesamiento de datos en tiempo real mejorará la calidad y la puntualidad de la información. Si bien hay muchas soluciones de EDW basadas en la nube para elegir, la implementación de estas soluciones puede ser bastante diferente. La elección de la solución de EDW correcta requiere planificación en torno a una estrategia de nube de datos eficaz y una comprensión no solo de la tecnología subyacente, sino también de las capacidades operativas y los modelos de precios. Con las exigencias de la empresa de hacer un uso más eficiente de los recursos, mejorar la visibilidad y optimizar el valor proporcionado por cada dólar de gasto en la nube, es fundamental que las organizaciones elijan el EDW adecuado basado en la nube.

La solución: Plataforma de datos empresariales sin servidor Google Cloud BigQuery

BigQuery es una plataforma de datos empresarial basada en la nube, totalmente administrada y sin servidor que admite análisis de datos de escala de gigabyte a petabytes y más (consulte la Figura 2). Ofrece un análisis de alta velocidad de grandes conjuntos de datos, a la vez que reduce o elimina las inversiones en administradores de infraestructura o bases de datos en las instalaciones. BigQuery amplía o reduce la escala de uso de hardware para maximizar el rendimiento de cada consulta, agregando y eliminando recursos informáticos y de almacenamiento según sea necesario.

BigQuery está diseñado para agilizar el análisis y el almacenamiento de datos a gran escala y, al mismo tiempo, ofrece opciones flexibles y optimiza los datos y los recursos de procesamiento para ahorrar costos. Algunas de las ventajas específicas de BigQuery para empresas que trabajan con macrodatos incluyen:

- **Optimización de costos.** Predicción de costos con opciones de precios de almacenamiento y procesamiento transparentes y flexibles, y limitación de costos con cuotas de proyecto y de recursos de usuario. Las organizaciones pueden optar por pagar tasas de almacenamiento lógicas (comprimidas) o físicas (sin comprimir) según sus características de datos y beneficiarse de la organización de niveles automática de almacenamiento activo y a largo plazo para reducir automáticamente los costos. Para el procesamiento, los usuarios pueden optar por pagar precios flexibles a pedido (por terabyte procesado) o precios predecibles de las ediciones BigQuery a fin de pagar solo por el tiempo de procesamiento consumido por las consultas, lo que les permite aprovechar los descuentos de escalamiento automático y de compromiso.
- **Tiempo de generación de valor.** Obtenga el entorno del almacén de datos en línea de manera rápida, sencilla y sin habilidades de administración de sistemas de nivel experto y bases de datos mediante la eliminación de la infraestructura y la reducción de la administración; esto se conoce como “No Ops” o “Zero Ops”. La evaluación y los servicios de migración de BigQuery, la transferencia de datos gratuita, la traducción automática de SQL y las tecnologías de motor de transmisión minimizan el tiempo, el costo y el esfuerzo para obtener las primeras informaciones de los datos.
- **Simplicidad.** Complete todas las tareas principales relacionadas con el análisis del almacén de datos a través de una interfaz intuitiva sin la molestia de administrar la infraestructura o los clústeres, o bien, sin tener que realizar funciones de vaciado y optimización que requieren mucho tiempo. Además, las herramientas gratuitas, como el servicio de migración de BigQuery, ayudan con cada fase de migración, incluidas la evaluación y planificación para la ejecución y verificación, y Looker Studio proporciona una visualización gratuita del uso y las tendencias de los datos. Google Cloud gratuito también ofrece herramientas como Dataform, Dataplex, Data Catalog, Hojas conectadas y BI Engine para facilitar la administración de consultas, canalizaciones de datos, metadatos y almacenamiento en caché para los usuarios y la organización.
- **Escalabilidad.** Amplía de manera transparente la capacidad de procesamiento para satisfacer las necesidades de operaciones de cualquier tamaño sin necesidad de planificación. Comience con una pequeña cantidad de datos y aumente la escala a exabytes según el tamaño, el rendimiento y los requisitos de costos. Las reservas con escala automática de BigQuery permiten a las organizaciones optimizar el costo y el rendimiento mediante la configuración de límites de escala automática para escalar automáticamente las ranuras a la demanda de la carga de trabajo cuando sea necesario y evitar el pago excesivo cuando no se requieren ranuras.

- **Agilidad.** Actúe con mayor rapidez cuando se presenten nuevas oportunidades de negocios, explore conjuntos de datos para descubrir nueva información, aproveche las capacidades de AA incorporadas de BigQuery ML (BQML) mediante consultas en SQL estándar e innove más allá de las expectativas tradicionales de un almacén de datos. BigQuery Omni proporciona una estrategia de análisis ágil de nubes múltiples sin tener que mover o copiar datos entre las nubes de Google, AWS y Azure. Analice datos de forma nativa desde una amplia variedad de almacenamiento, como Cloud SQL, Cloud Storage, Amazon S3, Azure Blob Storage, Google Drive, y un lakehouse compatible con hojas de cálculo de Google.
- **Velocidad.** Transfiera, consulte y exporte conjuntos de datos de tamaño de gigabytes a petabytes con velocidades impresionantes utilizando Google Cloud Platform como la infraestructura de nube subyacente.
- **Confiabilidad.** Garantice la disponibilidad y el tiempo de actividad constante de la plataforma de Google Cloud con replicación geográfica en los centros de datos de Google y un acuerdo de nivel de servicio (SLA) líder de hasta un 99,99 % de tiempo de actividad, en comparación con el 99,9 % para la mayoría de los otros proveedores.
- **Seguridad.** Proteja y controle el acceso a proyectos cifrados y conjuntos de datos a través de la administración de acceso e identidad en toda la nube (IAM) de Google y la opción de claves de cifrado administradas por el cliente (CMEK) para mantener el control de los datos cifrados.

Figura 2. Google Cloud BigQuery



Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

BigQuery escala automáticamente; identifica los requisitos de recursos para que cada consulta termine de manera rápida y eficiente y proporciona esos recursos para satisfacer la demanda. Otros departamentos o divisiones dentro de una organización pueden utilizar los recursos y la capacidad inactivos. Una vez que la carga de trabajo se completa, BigQuery libera o reasigna automáticamente los recursos para que estén disponibles para otros proyectos y otros usuarios. Mientras se transfieren datos y se procesan para obtener resultados, BigQuery ofrece grandes velocidades, incluso a escala de petabyte.

Para una mayor durabilidad de los datos, BigQuery proporciona alta disponibilidad y confiabilidad a través de la replicación geográfica que es completamente transparente para sus usuarios, sin el requisito de obtener los recursos físicos y el espacio para alojarlos. BigQuery proporciona copias de seguridad automáticas e instantáneas de conjuntos de datos, lo que permite una restauración de un punto en el tiempo y garantiza que se integren controles de servicio de nube privada virtual (VPC) y cifrado de datos con la opción de CMEK (tanto en tránsito como en reposo) para proporcionar niveles mejorados de seguridad.

En última instancia, BigQuery permite a las organizaciones abordar los desafíos de costo y complejidad asociados con la construcción y el mantenimiento de una infraestructura de datos a gran escala, escalable y resistente. Al aprovechar el enfoque basado en la nube de BigQuery, el tiempo y el costo tradicionalmente dedicados a proteger los datos y garantizar el tiempo de actividad casi se elimina. Con el manejo de la escalabilidad, replicación, protección y recuperación de Google Cloud, las organizaciones pueden centrarse más en obtener información valiosa, en lugar de la administración de infraestructura.

BigQuery proporciona una plataforma unificada para un amplio procesamiento de datos, transmisión, gobierno y capacidades de IA/AA. Al admitir varios formatos de datos (incluidos los estructurados y no estructurados), software de código abierto y servicios de socios, permite la ejecución de un ciclo de vida completo de los datos. Los clientes de BigQuery aprovechan su flexibilidad mediante el uso de varios idiomas de codificación, como SQL, Python, Spark y Natural Language, además del análisis entre nubes. Además, las funciones generativas de IA (GenAI) integradas de BigQuery permiten acelerar el tiempo de obtención de información mediante la creación y finalización de códigos, así como también mediante un chat en la consola. BigQuery ML permite a las organizaciones desbloquear los casos de uso de GenAI para datos estructurados y no estructurados mediante SQL simple.

BigQuery frente a servicios de EDW alternativos basados en la nube

ESG comparó la solución de EDW sin servidor de BigQuery con algunos servicios de EDW alternativos ofrecidos por AWS, Snowflake y Azure/Databricks. Si bien todas estas ofertas ofrecen ahorros significativos en costos, reducción de la complejidad y mayor agilidad comercial en comparación con una solución de EDW local, existen algunas diferencias importantes entre las ofertas (consulte la Figura 3).

Amazon Redshift. Redshift es el servicio de almacén de datos completamente administrado y basado en la nube de AWS. Al igual que un clúster local, Redshift se basa en el concepto de nodos (nodos virtuales) que se deben implementar, configurar y administrar. La administración se simplifica en gran medida, ya que no hay hardware que administrar y mantener físicamente. Para ampliar la implementación, se agregan simultáneamente nodos similares de una capacidad de almacenamiento y procesamiento fijos, lo que a veces da como resultado un mayor aprovisionamiento de capacidad de procesamiento y de almacenamiento para cumplir con los requisitos del otro. Los clústeres de Redshift requieren planificación y dimensionamiento, configuración y acceso a través de un “nodo líder” y actualizaciones programadas varias veces al año. Si bien AWS ofrece muchas opciones de precios flexibles, para lograr los mayores ahorros las organizaciones deben pagar por tres años de servicio por adelantado, vincularlos tanto a la solución del proveedor como al tipo de instancia particular que han elegido. AWS ha lanzado recientemente una versión sin servidor de Redshift, pero en el momento de la publicación, no hemos podido obtener suficiente información relacionada con el dimensionamiento y la administración para incluirla de manera creíble en este informe.

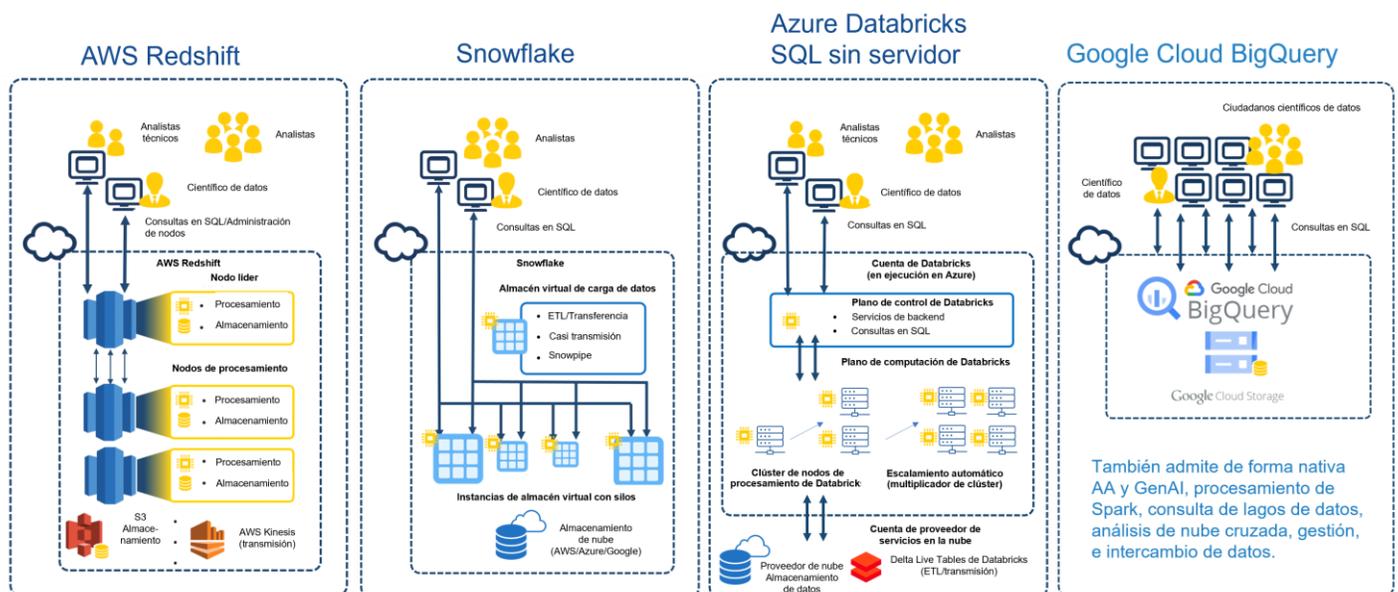
Snowflake. Snowflake es un almacén de datos administrado como un servicio (Data Warehouse as a Service, DWaaS) que se puede implementar en la infraestructura de Google Cloud, AWS o Azure. Snowflake no requiere la administración de hardware físico o virtual ni instalación de software ni mantenimiento. Snowflake también separa los recursos de procesamiento y de almacenamiento y utiliza una arquitectura de procesamiento paralelo masivo (Massively Parallel Processing, MPP) en segundo plano. En el momento de la implementación, los clientes deben seleccionar desde almacenes de datos virtuales preconfigurados en varios tamaños (pequeño, mediano, grande, extra grande, etc.), los precios se basan en las “horas de crédito” consumidas. El costo por crédito varía según la región, el servicio en la nube y el nivel de soporte. Los almacenes virtuales generalmente se colocan en silos según las unidades de negocios y se pueden pausar y reanudar automáticamente para limitar el costo y el tamaño cuando sea necesario. Snowflake también sugiere hacer uso de un almacén virtual dedicado para manejar la carga y transmisión de datos a fin de evitar la contención en los recursos necesarios para las consultas, y ofrece servicios alternativos con costos adicionales y ventajas y desventajas de confiabilidad.

Procesamiento sin servidor de Azure Databricks SQL. Esta es una versión de procesamiento sin servidor administrado de la arquitectura de la plataforma Databricks basada en Apache Spark, que se ejecuta en clústeres de procesamiento con una cuenta de Databricks en la infraestructura proporcionada por Azure. El plano de control de Databricks proporciona una aplicación web para administrar consultas en SQL, historial y extremos, que luego envían consultas a almacenes SQL sin servidor que se ejecutan en el plano de procesamiento sin servidor a través de una red privada mediante instancias de procesamiento proporcionadas en la cuenta Databricks. Los clientes

también pagan el almacenamiento de nube necesario para almacenar el lakehouse de datos y los datos que consume el sistema de archivos de Databricks (DBFS) o para almacenar los resultados de las consultas. El lakehouse de Databricks se basa en el formato de tabla de Delta Lake, que permite el uso de Databricks para el procesamiento de datos en tiempo real y la generación de informes de inteligencia del negocio (Business Intelligence, BI).

EDW de BigQuery sin servidor. La solución de BigQuery de Google Cloud es completamente sin servidor, con escalamiento automático, automantenimiento y ajuste automático. No hay nodos ni almacenes virtuales para planificar, configurar o escalar. Google maneja la complejidad del dimensionamiento, la administración y el mantenimiento de la infraestructura física tras bambalinas, lo que elimina la carga de la organización. Los usuarios finales obtienen la ventaja de que todos los recursos optimizados y ajustados automáticamente funcionan simultáneamente. Los clientes pueden pagar por la cantidad total de datos procesados por mes u optar por guardar con ediciones predecibles de BigQuery de pago por uso en función de la cantidad de “ranuras” requeridas (efectivamente, la cantidad reservada de recursos paralelos que se ponen a disposición para ejecutar consultas). Si bien los compromisos de ranura se pueden asignar a un departamento en particular para procesar consultas, las ranuras no utilizadas se asignan automáticamente a otros departamentos mediante la redistribución de la capacidad de inactividad y se ponen a disposición para realizar automáticamente el cálculo de las ráfagas, lo que ayuda a optimizar la utilización de ranuras. BigQuery también es compatible con el aprendizaje automático basado en SQL nativo, análisis de sistemas de información geográfica (Geographic Information Systems, GIS), tablas de BigLake para consultar lagos de datos y BigQuery Omni para análisis entre nube, lo que permite que todos los usuarios se beneficien de la potencia de los análisis avanzados sin la necesidad de integrarse con otras soluciones ni mover datos entre tecnologías y nubes.

Figura 3. Comparación funcional de soluciones de EDW basadas en la nube



Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

Validación económica de Enterprise Strategy Group

El proceso de validación económica de Enterprise Strategy Group (ESG) es un método comprobado para comprender, validar, cuantificar y modelar las propuestas de valor económico de un producto o solución. El proceso aprovecha las competencias principales de ESG en el análisis del mercado y de la industria, la investigación prospectiva y la validación técnica o económica. Para validar las suposiciones y los costos incluidos en el análisis, ESG realizó entrevistas detalladas y revisó estudios de casos de organizaciones que habían migrado previamente sus operaciones de una solución de EDW heredada en las instalaciones a BigQuery. Esto nos ayudó a comprender y cuantificar mejor cómo el cambio a BigQuery impactó o afectó a sus organizaciones. Utilizamos nuestros hallazgos para crear un modelo económico detallado que comparara los costos y beneficios esperados de BigQuery con las soluciones basadas en la nube de la competencia.

Descripción general del valor económico: BigQuery frente a soluciones de EDW alternativas basadas en la nube

El análisis económico de ESG reveló que BigQuery puede proporcionar importantes ahorros operativos y de capital, así como beneficios tangibles en comparación con otras soluciones de EDW basadas en la nube. ESG descubrió que BigQuery les proporcionó a los clientes ahorros y beneficios significativos en estas categorías:

- **Eliminación de la inversión inicial y la planificación.** El diseño sin servidor de BigQuery se factura mensualmente con opciones flexibles y predecibles de precios de procesamiento y almacenamiento que pueden adaptarse mejor a la necesidad de diferentes organizaciones y cargas de trabajo. Estas opciones eliminan la necesidad de pagar meses o incluso años antes para reducir el gasto en la nube y elimina las conjeturas y la planificación necesarias para ajustar el tamaño de los recursos. Esto proporciona a la organización la mayor agilidad empresarial cuando se trata de los costos de los servicios de nube.
- **Reducción de los gastos operativos.** BigQuery elimina la necesidad de administrar nodos de EDW virtuales, así como la necesidad de monitorear, solucionar problemas, actualizar, ajustar y planificar el crecimiento. BigQuery Storage se optimiza automáticamente para el costo, los parches y el mantenimiento no son necesarios, y el equipo de soporte está bien capacitado y responde. Esto deja a los administradores más tiempo para centrarse en otras áreas de la empresa.
- **Mayor agilidad comercial y reducción del costo de la administración diaria.** BigQuery aumenta o disminuye la escala según sea necesario para satisfacer las cambiantes demandas comerciales, lo que permite a las organizaciones actuar rápidamente sobre nuevas oportunidades sin necesidad de planificar los requisitos de configuración, pausar las bases de datos o iniciar instancias dedicadas de almacenes para cada organización. La solución también ayuda a eliminar o reducir el tiempo dedicado a la administración de la base de datos, a extraer, transformar y cargar (ETL) la administración y a la modificación de nuevos esquemas. BigQuery proporciona capacidades nativas de IA/AA e integración fácil para GenAI. También es compatible con la consulta nativa de lagos de datos con BigLake y una estrategia de análisis de varias nubes con BigQuery Omni.
- **Menor costo y complejidad de ejecutar cargas de trabajo de AA y casos de uso de GenAI.** El motor de inferencia BigQuery ML permite el uso de SQL simple directamente dentro de BigQuery para ejecutar predicciones mediante modelos alojados de forma remota y modelos de base de Google para los datos estructurados en almacenamiento administrado por BigQuery y datos no estructurados en Google Cloud Storage. Esto elimina la necesidad de movimiento de datos, lo que reduce la complejidad, los costos y los riesgos de seguridad para los clientes, a la vez que reduce la necesidad de depender de recursos especializados con conocimiento de lenguajes de programación avanzados.

Eliminación de la inversión inicial y la planificación del tamaño de implementación

La migración desde un EDW local a cualquier solución EDW basada en la nube elimina la necesidad de realizar una gran inversión inicial en infraestructura física y ayuda a evitar el aprovisionamiento en exceso de hardware de almacenamiento y procesamiento físico para el crecimiento y los aumentos de demanda. Sin embargo, no todas las soluciones EDW basadas en la nube evitan completamente la necesidad de realizar inversiones por adelantado, la necesidad de planificar el tamaño de la implementación o la necesidad de aprovisionamiento en exceso para manejar los aumentos de la carga de trabajo (consulte la Tabla 1).

- Eliminación de la inversión inicial.** BigQuery no requiere una inversión inicial y pagada para lograr el mayor nivel de ahorro para los recursos de procesamiento. Los descuentos ofrecidos a través de los compromisos de ranura de las ediciones de BigQuery son parte de un contrato, con el costo pagado mensualmente. Snowflake ofrece descuentos para pagos por adelantado y costos de almacenamiento más bajos para aquellos que tienen capacidad de reserva, pero no requiere una inversión por adelantado. (Sin embargo, el cliente debe pagar el almacenamiento no utilizado). AWS Redshift requiere contratos anuales e inversiones por adelantado para lograr el mayor nivel de ahorro. Este dinero debe pagarse con uno o tres años de anticipación, aunque el valor del servicio no se realice hasta meses o años después. De manera similar, Azure Databricks SQL Serverless ofrece descuentos por unidades Databricks Commit (DBCUs) de compra previa como una sola compra con uno o tres años de anticipación. Al igual que cualquier inversión de capital, las organizaciones deben asegurarse de tener en cuenta el costo adicional del capital a lo largo del tiempo al comparar las inversiones iniciales con los costos de compromiso a pedido mensuales.
- No hay planificación del tamaño de implementación de la infraestructura.** Con BigQuery, los clientes no tienen que planificar y ajustar el tamaño de la configuración para satisfacer sus necesidades. BigQuery es completamente sin servidor y los clientes no tienen que planificar ni ajustar la infraestructura de soporte. Con AWS Redshift, el tamaño y la cantidad de las instancias necesarias para manejar la transferencia y las cargas de trabajo deben predecirse y dimensionarse como si fueran nodos locales. Con Azure Databricks SQL sin servidor y Snowflake, los usuarios deben determinar la combinación de almacenamiento y el tamaño de la “camiseta” del grupo de nodos de procesamiento o el almacén de datos virtual que mejor se adapte a las necesidades de cada unidad de negocios. Snowflake también recomienda un almacén dedicado que se utiliza para cargar datos.
- No es necesario aumentar la implementación.** BigQuery a pedido no requiere el monitoreo de la utilización de la capacidad ni el crecimiento del entorno cuando se alcanzan los límites, y los compromisos de ranura de ediciones se pueden ampliar fácilmente cuando sea necesario. Las ediciones de BQ aprovechan el escalamiento automático para aumentar la escalar hasta un número máximo de ranuras predefinidas durante tiempos de ráfaga y luego vuelven a los niveles normales cuando no es necesario. Las implementaciones de AWS Redshift se pueden escalar, pero esto no solo requiere sobrecarga de administración, sino también cierto tiempo de inactividad temporal del clúster a medida que se agregan las instancias. Escalar los recursos para Snowflake y Azure Databricks SQL sin servidor es relativamente fácil, pero requieren cierto grado de planificación y actividad administrativa. Estas ofertas también ofrecen escalamiento automático de almacenes, pero estas ofertas deben escalarse en la granularidad del tamaño de clúster original (múltiplos) y requerir un mayor tiempo para implementar, calentar y limpiar después de su uso, lo que le cueste más al usuario. La Figura 4 ilustra la ventaja del escalamiento automático basado en ranuras para satisfacer mejor las necesidades de cambiar las cargas de trabajo desde una perspectiva de agilidad y ahorro de costos. Google Cloud estima que con el escalamiento automático granular, algunas organizaciones pueden reducir su gasto de procesamiento comprometido en un 30 % a un 40 %.

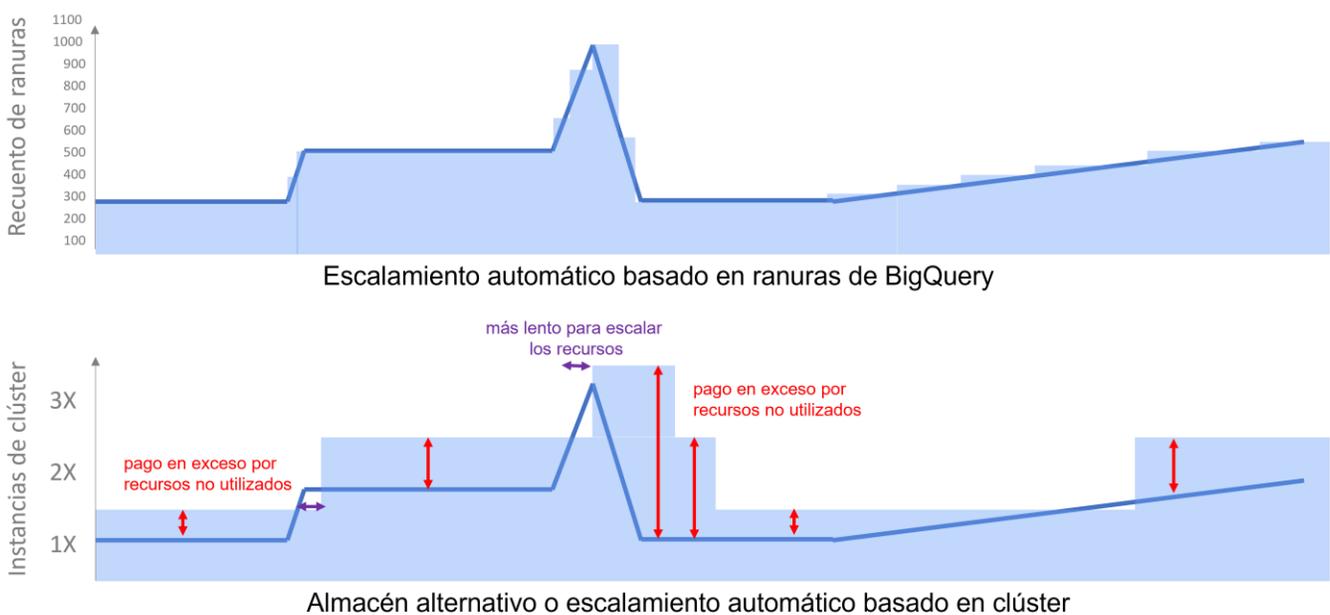
“Con Google, no estamos bloqueados en ninguna duración ni debemos pagar por adelantado para ahorrar como otros contratos en la nube, lo que es importante para nosotros a medida que aumentan nuestras capacidades y nuestras necesidades cambian. Puedo cambiar las reservas y los modelos de precios cada mes según lo necesite. Este modelo nos brinda mucha más flexibilidad que cualquier otra cosa que haya utilizado”.

“Tan pronto como implementamos los precios de las reservas de BigQuery, comenzamos a ver un cambio drástico en la facturación. Sin realizar otros cambios, hemos disminuido nuestro gasto mensual en un 30 %”.

La infraestructura BigQuery utiliza todos los recursos disponibles para procesar consultas y no tiene que ser aprovisionada en exceso para manejar aumentos inesperados en la carga de trabajo. Los clústeres de procesamiento de AWS Redshift, Snowflake Virtual Warehouses y Azure Databricks SQL sin servidor definen un conjunto fijo de recursos que están disponibles para que las consultas se ejecuten y también para manejar cualquier proceso de

transferencia, reagrupación y vaciado. En ese sistema, una implementación de configuración que es demasiado pequeña afecta negativamente las consultas, lo que lleva a las organizaciones a dimensionar y pagar las implementaciones de línea base para manejar el peor de los casos para cargas de trabajo por ráfagas. BigQuery permite que cualquier recurso de cualquier departamento de la organización utilice ranuras fijas y capacidad compartida disponibles, mientras que los departamentos de otras soluciones no pueden beneficiarse del procesamiento o el almacenamiento en estado inactivo o el aprovisionamiento en exceso en clústeres de otros departamentos, lo que limita aún más la eficiencia de los recursos según la cantidad de clústeres implementados. Aunque las implementaciones agrupadas en varios grupos en Azure Databricks SQL sin servidor y Snowflake pueden ayudar a superar las limitaciones de escalamiento, deben crecer en un mínimo de 2 veces los recursos, lo que contribuye aún más a la ineficiencia del aprovisionamiento en exceso en muchos casos y se limita a un número máximo de clústeres.

Figura 4. Escalamiento automático basado en ranura frente a basado en clúster



Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

- Costo reducido para migrar.** Si bien cada proveedor ofrece herramientas y servicios para ayudar a migrar datos a su base de datos en la nube, los clientes informaron que la migración a BigQuery era notablemente rápida y simple, con herramientas gratuitas proporcionadas por Google Cloud, mientras que otras herramientas de proveedores y servicios adicionales costaron más. Google Cloud ha realizado importantes inversiones en adquisiciones, mejoras y herramientas tecnológicas para la migración e incorporación. La herramienta de evaluación automatizada de Google ayuda a comprender los activos y priorizar las tareas. La transferencia de datos por lotes es gratuita. La traducción interactiva de SQL y por lotes a través de la adquisición de CompilerWorks reduce el tiempo y el esfuerzo manuales y reduce el riesgo de errores. La herramienta de validación de datos de Google Cloud valida las tablas de origen y destino. Al reducir los recursos necesarios para planificar, traducir, transferir, validar y capacitar a los clientes en un nuevo esquema, BigQuery puede limitar el costo operativo inicial y acelerar el tiempo hasta el valor comercial.

Tabla 1. Comparación de soluciones de EDW basadas en la nube: Inversión inicial, planificación y agilidad

| | Google BigQuery | AWS Redshift | Snowflake | Azure Databricks SQL sin servidor |
|---|---|--|---|---|
| Inversión inicial (Pagado antes de la prestación de servicios) | Ninguna. Los contratos de tarifa plana aún se pagan mensualmente, no por adelantado. | Contratos de un año o tres años con inversiones iniciales necesarias para recibir tasas competitivas. | Ninguna para procesamiento. Capacidad de almacenamiento reservada necesaria para el precio más bajo, ya sea que se utilice o no. | Pago por uso con una opción para descuentos por compromiso de uso (almacenamiento). Se requiere la compra de crédito por adelantado (DBCU) para obtener los mejores ahorros. |
| Dimensionamiento y planificación | No se requiere planificación. BigQuery maneja todas las asignaciones de ranuras. No es necesario administrar el procesamiento. | Debe predecir los tamaños de instancia correctos y las instancias reservadas limitan la agilidad. | Es necesaria la predicción del tamaño de almacén virtual necesario para cada unidad de negocios. | Es necesaria la predicción del tamaño del clúster sin servidor y encenderlo manualmente. Se deben cancelar los clústeres inactivos o generarán costos (se pueden cancelar automáticamente). |
| Agilidad/ Crecimiento | Escalamiento automático de impacto cero. No se requiere ajuste para los precios a pedido. Puede establecer límites simultáneamente (valor predeterminado = 100). Se pueden aumentar ranuras adicionales cuando sea necesario. | El cambio de tamaño del clúster requiere tiempo de inactividad del clúster. Limitado a 15 consultas concurrentes por clúster. Puede escalar automáticamente a 10 clústeres, pero se limita a 50 consultas en cola. | Se escala agregando nuevos almacenes virtuales o aumentando el tamaño de los almacenes virtuales individuales. Los almacenes agrupados en varios grupos deben crecer por dos. Límite de 8 consultas concurrentes por almacén, escala automática a 10 almacenes (80 consultas concurrentes). | La escala automática se realiza agregando múltiplos de clústeres de nodos de tamaño seleccionado. Limitado a 10 consultas concurrentes por clúster. Toma tiempo para escalar y calentar nuevas copias de clústeres. |
| Aprovisionamiento en exceso para aumentos de rendimiento o crecimiento de la capacidad | Ninguno. | Se deben aprovisionar en exceso suficientes instancias para manejar el peor de los casos o aumentar la sobrecarga operativa de almacenes y el tiempo de inactividad de escalamiento hacia arriba y hacia abajo, según sea necesario. | Se deben aprovisionar en exceso instancias para manejar el peor de los casos o aumentar la sobrecarga operativa de almacenes de escalamiento hacia arriba y hacia abajo, según sea necesario. | Admite escalamiento automático, lo que limita la necesidad de aprovisionamiento en exceso, pero se necesita la predicción de nodos de trabajador mínimos y máximos para cada trabajo. |
| Costos de almacenamiento | Elección de almacenamiento físico (comprimido) o lógico para optimizar el costo. | El almacenamiento y el procesamiento separados solo están disponibles en nodos RA3 con almacenamiento administrado. Los tipos de nodos más antiguos se deben configurar con capacidad de almacenamiento SSD o HDD interna. | Almacenamiento microparticionado con capacidad de datos de límite de compresión escaneada. El servicio de optimización de búsqueda puede preindexar algunas consultas a un costo adicional. | Costo de almacenamiento pagado a Azure. Parquet y Delta Lake comprimen el almacenamiento para reducir los costos. |

Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

Reducción de los gastos operativos

La eliminación de la necesidad de administrar y mantener el hardware en las instalaciones genera ahorros operativos importantes para las organizaciones que migran a los servicios basados en la nube. La naturaleza completamente sin servidor de BigQuery trae consigo ventajas operativas adicionales, incluso cuando se compara con otros servicios de EDW basados en la nube destinados a reducir estos costos (consulte la Tabla 2).

- **Eliminación de la administración de nodos.** BigQuery es completamente sin servidor y sin estado. AWS Redshift requiere la administración de nodos virtuales para tareas como configurar el acceso a la red y los VPN, asignar recursos, actualizar software, mantener el acceso a los recursos y aumentar y reducir el clúster. Snowflake y Azure Databricks SQL sin servidor son más simples que Redshift, pero aún requieren escalamiento y administración manuales para varios almacenes virtuales o clústeres de procesamiento. Un cliente dijo: “Teníamos aproximadamente 60 personas planificando, administrando y manteniendo nodos de infraestructura física o virtual en las instalaciones y en AWS. Esa fue una parte importante de nuestro equipo que pudimos redistribuir en otras actividades”.
- **Administración de almacenamiento reducida.** BigQuery es la única solución de EDW basada en la nube de las cuatro que ofrece almacenamiento totalmente administrado y autooptimizable. Los datos de más de 90 días se mueven automáticamente al almacenamiento rentable para reducir los costos en un 50 %. BigQuery mantiene copias replicadas de datos para la disponibilidad de 11 nueves, cifra todos los datos antes de escribir en el disco, admite clones e instantáneas eficientes basados en delta y permite reparticionar las tablas.
- **Sin mantenimiento de procesamiento.** BigQuery, Snowflake y Azure Databricks SQL sin servidor se mantienen sin interrupciones para el usuario en segundo plano. Los nodos de Snowflake mantienen el estado, y Azure Databricks SQL sin servidor y Redshift tienen cachés SSD para datos, por lo que llevarlos hacia abajo puede afectar el rendimiento. AWS Redshift requiere que se programen las ventanas de mantenimiento manual y las actualizaciones pueden producir tiempo de inactividad.
- **Estrategia unificada de nube de datos.** Las organizaciones modernas se benefician del uso de la estrategia unificada de nube de datos de Google Cloud para aprovechar al máximo sus datos y reducir aún más el costo y la complejidad mediante la integración y optimización de BigQuery con otros productos de datos de Google Cloud. Google Cloud BigLake permite a las organizaciones unificar los almacenes y lagos de datos en un solo conjunto de datos. Esto simplifica la gestión de datos y la gestión del ciclo de vida y reduce las copias duplicadas de los datos, a la vez que simplifica y acelera la transferencia y la transformación de datos. BigQuery Omni permite a los usuarios acceder a los datos en otras nubes públicas, y BigSearch proporciona la capacidad de consulta para el análisis de registro basado en JSON.

“Nuestro principal dolor de cabeza con nuestro almacén de datos en la nube anterior era mantener la base de datos y ajustar y configurar el clúster. Con BigQuery podemos enfocarnos en nuestras carteras, los datos y nuestros resultados”.

Tabla 2. Comparación de soluciones de EDW basadas en la nube: Gastos operativos

| | Google BigQuery | AWS Redshift | Snowflake | Azure Databricks SQL sin servidor |
|--|---|--|--|---|
| Administración de nodos virtuales | Ninguno; completamente sin servidor. | Nodo de administración requerido con configuración manual y escalamiento de nodos. | Sin servidor, pero es posible que los usuarios deban administrar almacenes de datos en silos y se sugiere un almacén dedicado para cargar datos. | Sin clúster para administrar, pero los usuarios deben ajustar el tamaño y comenzar los clústeres, administrar el acceso de los usuarios, convertir los datos en formato delta (que requiere procesamiento adicional) y administrar las políticas de escalamiento/terminación. |
| Almacenamiento Administración | El almacenamiento autooptimizado transfiere automáticamente los datos al almacenamiento rentable. | Migración y envejecimiento del almacenamiento manual. | Debe administrar el movimiento de datos y decidir entre un almacenamiento costoso y rápido y un almacenamiento de capacidad económico. | Debe administrar las opciones de almacenamiento (almacenamiento de Azure Blob o Data Lake). El formato de almacenamiento requiere la optimización periódica del almacenamiento. |
| Mantenimiento de EDW/nodo | Administrado por Google Cloud en segundo plano sin tiempo de inactividad. | Actualizaciones manuales de nodos durante el tiempo de inactividad programado. | Administrado por Snowflake en segundo plano sin tiempo de inactividad. | Administrado por Databricks en segundo plano sin tiempo de inactividad. |

Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

Mejora de la agilidad comercial y reducción del costo de la administración diaria de EDW

La administración diaria de EDW a menudo consta de tareas como la administración del proceso ETL entre sistemas, la configuración, la administración y el monitoreo de la plataforma y la base de datos, la solución de problemas, el control del acceso, el mantenimiento de la seguridad, el desarrollo de nuevos servicios; la colaboración con analistas comerciales, la provisión de datos para informes y paneles, y la integración de aplicaciones y otros servicios de nube. ESG descubrió que, cuando se compara con una solución local, BigQuery y las otras soluciones basadas en la nube pueden eliminar algunas tareas en conjunto y hacer que otras tareas sean más simples y rápidas a través de la automatización e integración. BigQuery ofrece simplicidad de administración adicional, incluso cuando se compara con otros servicios basados en la nube (consulte la Tabla 3).

- **Se eliminó la necesidad de monitorear y cambiar el tamaño de los recursos virtuales.** Los administradores de BigQuery no necesitan pasar tiempo supervisando o escalando los recursos virtuales del sistema (nodos, unidades de procesamiento o almacenes virtuales). Debido a que los límites de simultaneidad de BigQuery son blandos, se pueden aumentar fácilmente cuando sea necesario. Para aquellos que desean monitorear el entorno de BigQuery, los paneles de monitoreo integrados permiten conocer el rendimiento de las consultas, las tendencias, la facturación y las auditorías del sistema, con la capacidad de crear alertas automatizadas basadas en los umbrales.
- **No es necesario realizar tareas periódicas de limpieza y optimización de EDW.** BigQuery realiza un reagrupamiento automático de tablas para un rendimiento óptimo y elimina la necesidad de realizar mantenimiento periódico, como procesos de vaciado que pueden consumir recursos y afectar negativamente el rendimiento. Algunas de las ofertas alternativas requieren procesos de reagrupación y vaciado manuales o automáticos que se ejecutan en segundo plano, y todas estas consumen recursos de procesamiento que se deben pagar, y las operaciones pueden afectar el rendimiento de las consultas activas.

- **Agilidad de negocios mejorada.** La escalabilidad con capacidad de respuesta significa que las consultas se pueden completar antes, que las decisiones informadas se puedan tomar más rápido y que los recursos para satisfacer las oportunidades comerciales con plazos críticos están disponibles en todo momento. Los recursos no tienen que asignarse ni expandirse para satisfacer las necesidades de nuevas oportunidades. Los clientes de BigQuery pueden compartir los compromisos de ranuras entre proyectos para manejar los cambios en la demanda de carga de trabajo. Los clientes de BigQuery informaron que eran mucho más ágiles y capaces de actuar en oportunidades mucho más rápido, sin impacto en el negocio.
- **ETL simplificado y gestión de fuentes de datos.** El ESG de los clientes habló con quien informó que las opciones flexibles les permiten organizar los datos en las instalaciones o en Google Cloud y almacenar los datos en una variedad de opciones dentro de Google Cloud: Cloud Spanner, BigTable, Cloud Storage, etc., o desde fuentes de datos externas. Estas opciones flexibles permiten mayores posibilidades y capacidades mediante el uso de otros productos de Google Cloud. Google Cloud Dataform ayuda a simplificar el diseño y la administración de la cartera de datos. Existen algunas opciones alternativas en el ecosistema de AWS, pero Snowflake y Azure Databricks SQL Serverless dependen de las integraciones con los servicios de nube pública.
- **Transferencia simplificada de datos.** El procesamiento de datos es una de las fortalezas de BigQuery cuando se compara con las soluciones de EDW alternativas basadas en la nube. ESG revisó los resultados de pruebas creíbles de terceros que mostraron que BigQuery obtuvo un puntaje de 1,9 veces a 2,4 veces mayor que las soluciones alternativas en las capacidades de transferencia de datos. La transferencia de datos por lotes es gratuita y consume cero capacidad de consulta del cliente en BigQuery. AWS Redshift y Snowflake requieren configuraciones y tareas mucho más complejas para administrar las operaciones de transferencia, y las operaciones de transferencia aprovechan los recursos de procesamiento de los usuarios, lo que esencialmente requiere el aprovisionamiento en exceso de recursos de procesamiento para satisfacer la demanda de la carga de trabajo. El servicio de Snowflake sugiere implementar un almacén virtual dedicado para la carga y la transferencia de datos, lo que agrega complejidad en cuanto a costos y administración. El servicio de Snowpipe de Snowflake puede facilitar la transferencia, pero agrega costo, y los trabajos consumen recursos de procesamiento adicionales.
- **Tareas simplificadas heredadas relacionadas con EDW.** Los clientes de BigQuery informaron un ahorro de tiempo al eliminar las copias de seguridad y las tareas de mantenimiento del sistema, simplificando los procesos como la incorporación, la administración de las prioridades de la carga de trabajo y el mantenimiento de las particiones.
- **Seguridad y control de acceso simplificados.** Google Cloud Deployment Manager puede ayudar a automatizar la creación de roles personalizados de IAM, lo que proporciona una administración simplificada de proyectos y control de acceso a través de la flexibilidad para administrar rápidamente el acceso a los recursos por función, organización, etapa del ciclo de vida de los datos o proyecto. BigQuery elimina la necesidad de administrar usuarios individuales u organizar y otorgar permisos, y tiene funciones de seguridad y cifrado incorporadas con claves administradas por el cliente o BigQuery.

“Descubrimos que Redshift fue engorroso de implementar y escalar, BigQuery ha sido mucho más fácil. Nos proporciona una mejor visión de nuestros proyectos y proporciona costos más predecibles para lo que estamos ejecutando”.

Tabla 3. Comparación de soluciones de EDW basadas en la nube: Mayor agilidad y administración simplificada de EDW

| | Google BigQuery | AWS Redshift | Snowflake | Azure Databricks SQL sin servidor |
|---|---|---|--|--|
| Planificar y ajustar el tamaño de los requisitos de recursos virtuales para nuevas cargas de trabajo | No se requiere tamaño adicional. | Instancias de tamaño AWS, considerando el procesamiento, la memoria y el almacenamiento local. | Ajustar el tamaño del almacén virtual en compartimentos o escale el almacén existente. | Predecir los tipos de instancia de IaaS de valor inicial y el tamaño inicial del grupo. |
| Aprovisionamiento de clúster/almacenes virtuales/redes | No se requiere aprovisionamiento. | Aprovisionamiento de recursos de VPC, red, acceso de usuario y almacenamiento en nuevas instancias. Tiempo de inactividad requerido. | Aprovisionamiento simple. | Aprovisionamiento simple. |
| Tareas periódicas de optimización y limpieza de EDW | Reagrupación automática de tablas en segundo plano. No es necesario realizar el vaciado de los procesos. No se factura por el procesamiento. | Optimización automática de la tabla. Operación de vaciado frecuente necesaria en tiempos que no son pico para limpiar las tablas y mantener un rendimiento constante. Facturado por el procesamiento consumido. | Reagrupamiento automático y operaciones de vaciado sin impacto en la producción. Facturado por el procesamiento utilizado. | El servicio administrado realiza la ejecución del proceso de compactación y vaciado para optimizar el rendimiento de almacenamiento y consultas. Facturado por el procesamiento utilizado. |
| Administración/monitoreo/escalamiento de recursos virtuales | No se requiere. | Monitoreo del clúster para comprobar el rendimiento y aumento/reducción de la escala según sea necesario. Las instancias reservadas no se pueden volver a escalar. | Monitoreo del clúster para comprobar el rendimiento y aumento/reducción de la escala. La dependencia del almacenamiento en caché puede dar como resultado un rendimiento impredecible. | Supervisión del desempeño del trabajo anterior y ajuste de los recursos de manera iterativa, o utilización de la función de escala automática. |
| Administración de los parches de acceso de usuario/seguridad | Controles IAM simplificados; acceso basado en roles con parches automáticos. | Administración de usuarios y seguridad para cada clúster. Aplicación de parches durante las ventanas programadas. | Permisos simplificados a nivel de recursos y basados en roles con aplicación automática de parches. | Administración de usuarios y seguridad para cada clúster. Aplicación automática de parches. |
| Administración de la carga/transferencia/transmisión de datos | Opciones flexibles de transferencia, transformación y transmisión de datos. El rendimiento de red más alto. ETL/ELT con Cloud Data Fusion. | La carga y la transmisión de datos complejos requieren el servicio Kinesis. La transferencia consume recursos de procesamiento asignados. ETL/ELT con AWS Glue. | Se sugiere que se ejecute un almacén separado para propósitos de carga de datos. El servicio Snowpipe está cerca de tiempo real. | Administración de grupos separados y cree trabajos para ejecutar carteras de datos y transformaciones. Transmisión con Delta Lakes y Kafka. |
| Integración con otras herramientas y servicios basados en la nube | Diseñado para la integración con muchos servicios de Google Cloud, incluidos DataFlow, DataProc, CloudDB, PubSub y Hojas de cálculo, que requieren una mínima sobrecarga de administración. Análisis de varias nubes con BQ Omni. | Integración manual con servicios AWS; algunos gastos generales de administración y configuración. | La integración empujada con los servicios de otros proveedores requiere administración y sobrecarga operativa. | Las integraciones limitadas con algunas herramientas de nube y de código abierto para BI, ETL y IDE requieren administración y sobrecarga operativa. |
| Integración con IA/AA | Capacidades incorporadas para crear y ejecutar modelos de AA. Compatibilidad con GenAI y casos de uso avanzado de AA a través de integraciones de Vertex AI. | Capacidades de AA a través de Sagemaker. GenAI con Amazon Q. | Capacidades de AA limitadas; debe utilizar otras soluciones. Varias herramientas GenAI en vista previa. | La oferta sin servidor no admite AA |

Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

Consideraciones de migración

La migración de cargas de trabajo de EDW en las instalaciones a una solución o servicio de EDW basado en la nube proporciona ventajas y oportunidades significativas para ahorrar costos, pero esto no se puede lograr de manera realista durante la noche. Las organizaciones deben analizar qué tareas están involucradas y diseñar un plan sólido para decidir cuándo trasladar su EDW a la nube.

- **Estrategia de migración para EDW.** Las organizaciones deben considerar si desean migrar datos de una sola vez y cambiar el conmutador de las operaciones existentes o migrar las operaciones lentamente durante un período de años mientras eliminan gradualmente el EDW local. Las organizaciones deben considerar los costos involucrados para migrar en términos de tiempo, transferencia física de datos, servicios profesionales, posible tiempo de inactividad, etc.
- **Costos de desarrollo y pruebas.** Las organizaciones deben considerar los costos involucrados en la creación de herramientas de aplicaciones, la transformación de datos en nuevos esquemas, la reescritura de consultas optimizadas, las pruebas, la validación, la solución de problemas, la creación de aplicaciones personalizadas, volver a capacitar a los desarrolladores, etc.
- **Rediseño del proceso.** El cambio presenta la oportunidad de mejorar muchos procesos y sistemas. Las organizaciones deben considerar el costo en términos de tiempo y dinero para realizar cambios en el proceso ETL, incluidas las capacidades de transmisión, la integración con otros productos en la nube y la educación de los equipos de negocios y los recursos de TI en el nuevo proceso y las nuevas herramientas.
- **Costos de servidor y software.** Las organizaciones deben considerar los costos asociados con los recursos complementarios, como los servidores o software de almacenamiento provisional en las instalaciones o basados en la nube y las licencias SaaS, sin olvidar equilibrar estos nuevos costos con cualquier ahorro potencial obtenido mediante la migración a la nube.
- **Estrategia de varias nubes.** Las organizaciones deben considerar la capacidad de aprovechar su almacén de datos basado en la nube para acceder a los datos almacenados en cualquier nube pública. BigQuery Omni permite que los datos se analicen mediante BigQuery en AWS o Azure sin tener que migrar o copiar datos a Google Cloud. Esto proporciona ahorro de costos y mucha mejor flexibilidad y agilidad.

Para ayudar a garantizar una migración exitosa, Google Cloud proporciona a las organizaciones un marco de migración prescrito con recursos, financiamiento y una metodología comprobada.²

Análisis de Enterprise Strategy Group

ESG aprovechó la información recopilada a través del material proporcionado por el proveedor, el conocimiento público y de la industria sobre economía y tecnologías, y los resultados de las entrevistas con los clientes para crear un modelo de tres años para el modelo de TCO/ROI que compara los costos y beneficios de satisfacer los requisitos de EDW de una organización modelada con BigQuery frente a otras tres soluciones EDW basadas en la nube. Las entrevistas de ESG con clientes que tienen experiencia con varias soluciones de EDW en la nube e información de precios disponible públicamente y pautas de dimensionamiento, combinadas con nuestra experiencia en modelado económico y validación técnica, ayudaron a formar la base de las suposiciones utilizadas en nuestro escenario modelado.

Para reflejar el tamaño y la escala de los clientes entrevistados, ESG consideró un escenario de una organización grande típica (más de 10 000 empleados) con requisitos para almacenar y analizar aproximadamente 200 TB de datos generados por varias unidades de negocio. De estos datos, se asumió que 110 TB era un conjunto de datos de tamaño fijo de tablas históricas y actualizadas que no vencieron, y ESG asumió que 250 GB de datos se transfirieron a diario con un período de retención de 365 días. Nuestros modelos predijeron que, con base en estos parámetros, cada una de las soluciones requeriría casi 100 TB de almacenamiento después de una compresión del 50 %. Asumimos que la organización ejecutó de manera coherente una línea de base promedio de consultas en toda la organización, lo que dio como resultado la necesidad de unas 800 ranuras BigQuery en todo momento. Además, hubo algunos períodos de actividad intensos durante el día en que la necesidad de ranuras adicionales aumentó en aproximadamente 1000 ranuras, durante un promedio de aproximadamente una hora por día, y, a veces, las consultas de alta prioridad ejecutaron hasta 2000 ranuras para períodos cortos y poco frecuentes. (Suponemos que esto promedió aproximadamente 5 minutos diarios).

² Fuente: Google Cloud, [Migrar Teradata y otros almacenes de datos a BigQuery](#), agosto de 2019.

ESG asumió que la implementación de BigQuery tenía un precio que utilizaba los precios de procesamiento de pago por uso de Enterprise Edition con una capacidad comprometida de un año de 800 ranuras y la capacidad de aumentar la escala automáticamente hasta 2000 ranuras adicionales para manejar las demandas máximas. Las otras soluciones se dimensionaron con estimaciones conservadoras de relaciones entre el número de ranuras fijas BigQuery y unidades de procesamiento elástico (ECU) (AWS Redshift) y un tamaño de implementación de clúster Azure Databricks SQL sin servidor y Snowflake equivalente (se muestra en la Tabla 4). Estas relaciones se obtuvieron de buena fe mediante el análisis de la información recopilada en el campo de pruebas preliminares de concepto, reemplazos de soluciones y datos publicados. Asumimos que la implementación de Redshift se ejecutaría de manera continua. También asumimos que tanto las configuraciones de Snowflake como las de Azure Databricks SQL sin servidor servirían las cargas de trabajo equivalentes a 800 ranuras de base, las 24 horas, los 7 días de la semana y, por lo tanto, se escalarían de manera automática para alcanzar los niveles de ranura más altos equivalentes, con un tiempo adicional en sobrecarga para acomodar un mínimo de suspensión automática de 10 minutos y completar los procesos de limpieza. En la práctica, las soluciones competitivas pueden requerir aún más recursos o tiempo de escalamiento automático para lidiar con la sobrecarga adicional de recursos necesaria para los procesos de transferencia, reagrupación y vaciado de datos, que no tienen costo alguno y no tienen un impacto negativo en el rendimiento de BigQuery. Para hacer que las soluciones sean lo más equivalentes posible desde una perspectiva de características hasta la edición empresarial de BigQuery, utilizamos los precios para la edición empresarial de Snowflake y el nivel de Azure Databricks SQL sin servidor Premium.

Las suposiciones de configuración para cada solución se muestran en la Tabla 4.

Tabla 4. Suposiciones de configuración utilizadas en la situación modelada de tres años de ESG

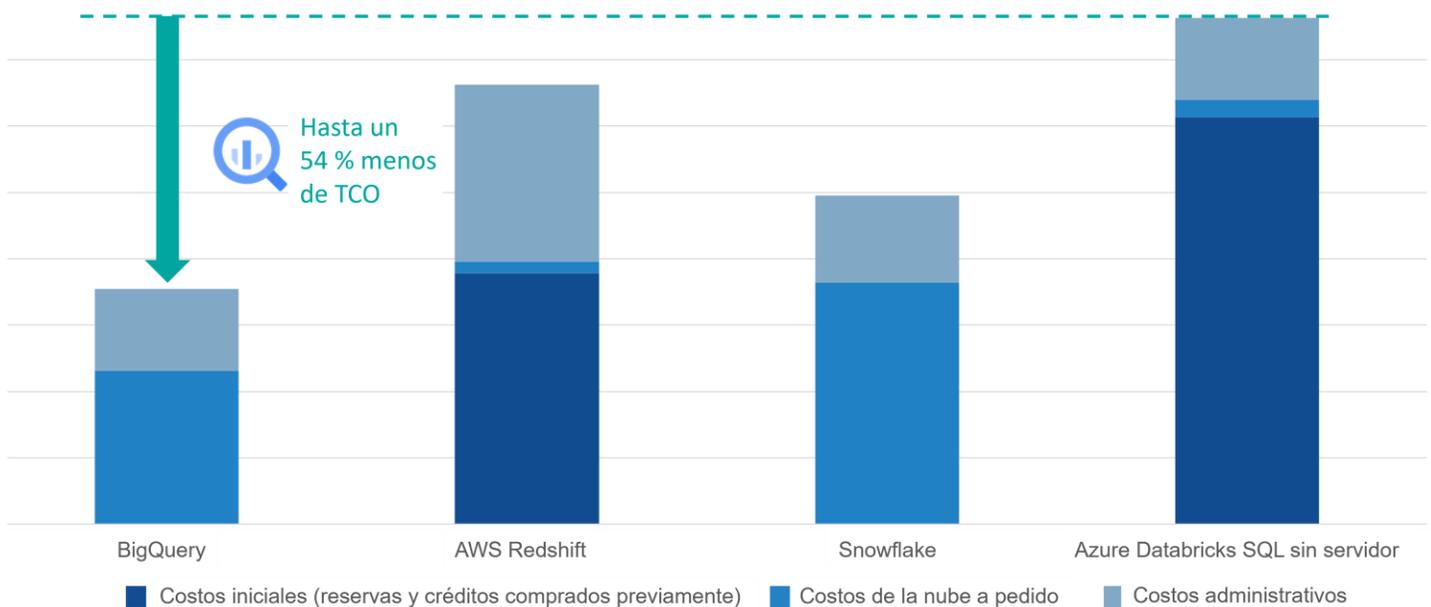
| | Google BigQuery | AWS Redshift | Snowflake | Azure Databricks SQL sin servidor |
|---|---|---|---|--|
| Opción de precios de procesamiento/servicio | Edición empresarial con un compromiso de un año (pagado mensualmente) | Precio inicial de instancia reservada por un año (ahorro del 34 %) (pagado por adelantado anualmente) | Enterprise Edition (AWS EE. UU. Este) (\$3.00/crédito) | Unidades DBCU sin servidor Azure Databricks SQL adquiridas previamente, con 1 año de anticipación, con un descuento del 23 %. (Nivel de prima) |
| Configuración de procesamiento/servicio | 800 ranuras de capacidad comprometida. Escala automática a 1000 ranuras adicionales para un promedio de 1 hora/día y a 2000 ranuras para un promedio de 5 minutos/día | 8 instancias ra3.16xgrandes | 18 créditos (24 x 7) 22 créditos (1,2 horas al día) 44 créditos (0,26 horas al día) | Instancia extragrande (80 DBU) (24 x 7) Escala automática a picos: Dos copias horizontales (promedio de 1,2 horas por día) Tres copias horizontales (promedio de 0,26 horas por día) |
| Costo anual de capital para gastos iniciales | N/A | 8 % | 8 % (solo almacenamiento) | 8 % |
| Almacenamiento de nube | 99,56 TB de almacenamiento físico (comprimido) (optimizado automáticamente para ahorrar costos) | 99,56 TB de almacenamiento administrado | 99,56 TB de almacenamiento de capacidad (pago inicial por un ahorro del 43 %) | 99,56 TB de almacenamiento Azure Blob |
| Servicio de transmisión/carga de datos | Inserciones de transmisión | AWS Kinesis | Snowpipe estimado como almacén de datos de gran tamaño (8 créditos x 8 horas/día) | Configuración de Azure Databricks para la transferencia y la transmisión (nivel estándar, 6 DBU) |

Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

Debido a que esto no se diseñó como un estudio de transferencia de cargas de trabajo en las instalaciones a la nube, ESG no estimó el tiempo y el costo de la migración y complejidad de la solución inicial para cada proveedor en este análisis.³ Asumimos que era igual para todas las soluciones. ESG también asumió que el rendimiento proporcionado por estas configuraciones satisfaría los acuerdos de nivel de servicio (Service Level Agreement, SLA) de la empresa y no se le otorgó ninguna ventaja de rendimiento directo a una solución por sobre la otra. En la práctica, estos factores pueden variar ampliamente según el entorno de una organización y deben ser considerados cuidadosamente por la organización. BigQuery puede, de hecho, tener ventajas en ambos aspectos, pero de manera conservadora no asignamos ninguna ventaja de costo a ninguna solución basada en el tiempo y la complejidad para migrar o mejorar/predecir el rendimiento.

Según nuestras suposiciones conservadoras, ESG descubrió que la solución BigQuery proporcionó un costo total de propiedad (Total Cost of Ownership, TCO) un 54 % menor que Azure Databricks SQL sin servidor, un TCO un 47 % menor que AWS Redshift y un TCO un 26 % menor que Snowflake durante el período de tres años modelado. La Figura 5 muestra el TCO estimado de tres años para cada una de las cuatro soluciones EDW basadas en la nube para satisfacer las necesidades del escenario modelado.

Figura 5. Costo total de propiedad de la solución del almacén de datos basado en la nube de tres años



Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

Gasto inicial en la nube (inversión de capital prepago)

Al estimar el gasto en la nube, comparamos de manera conservadora los precios de Enterprise Edition de Google Cloud con una solución rentable para cada una de las tres soluciones de la competencia durante el período de tres años. En el caso de AWS Redshift, esto implicó la firma de acuerdos y el pago anticipado de un año por adelantado para alcanzar el nivel informado de ahorros (ahorro del 34 %) para los recursos de procesamiento. Con el procesamiento de Azure Databricks SQL sin servidor, la compra de créditos DBCU con

³ ESG estudió el costo de mover cargas de trabajo en las instalaciones a la nube en [Las ventajas económicas de migrar cargas de trabajo del almacén de datos empresariales a Google BigQuery](#), marzo de 2019.

un año de anticipación produjo ahorros del 23 %. Para cualquier pago realizado con un año de anticipación, calculamos el impacto financiero de pagar el servicio antes de recibir el beneficio calculando un costo de capital conservador del 8 % a cualquier parte del pago restante en el que el beneficio aún no se haya obtenido (como un cálculo de pago de intereses en disminución). Sin embargo, se debe tener en cuenta que si se reserva el procesamiento por adelantado, la solución se vuelve mucho menos flexible, lo que dificulta aprovechar las tecnologías mejoradas o reducir la escala de la implementación para bajar los costos. Los descuentos para estas soluciones se pueden aumentar aún más si se reservan recursos o se compran créditos con una anticipación de hasta tres años. Sin embargo, esto reduce aún más la flexibilidad de la solución. Para nuestra situación, asumimos que el procesamiento de BigQuery y Snowflake no requería pago inicial.

Costos de la nube a pedido

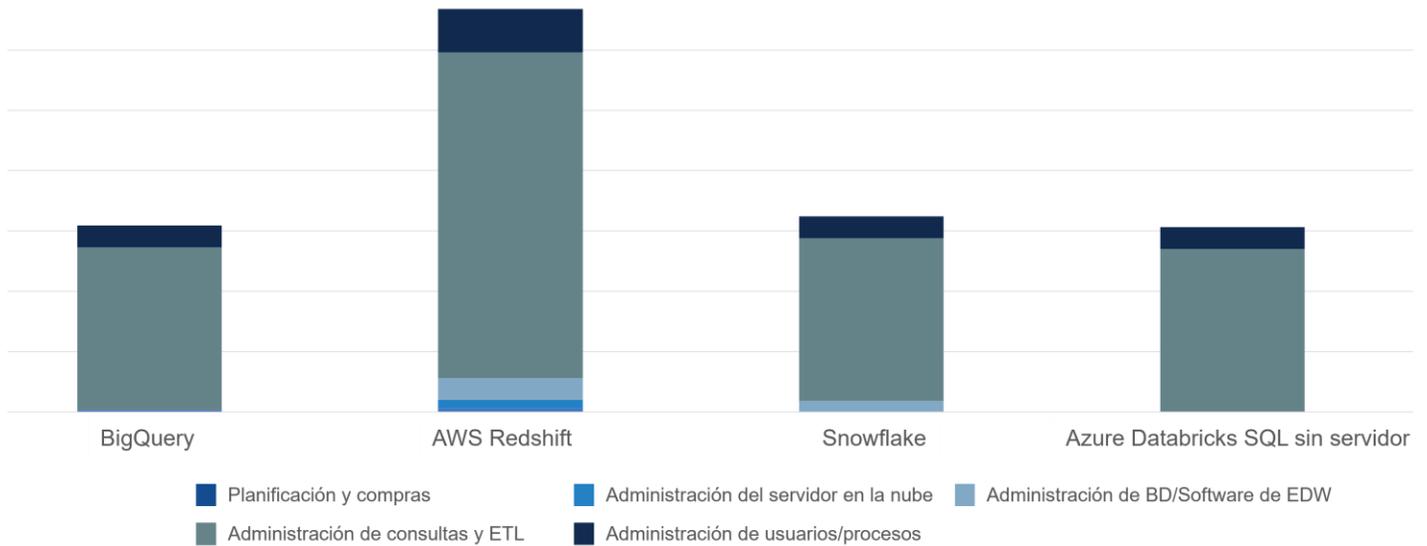
El gasto mensual en la nube constó de la factura estimada que recibiría la organización cada mes por procesamiento, almacenamiento o servicios. El costo de procesamiento de las instancias de AWS Redshift RA3 se pagó por adelantado anualmente (“costos iniciales” en la Figura 5). El gasto mensual en la nube (“Costos de nube a pedido” en la Figura 5) para AWS Redshift consistió en 99,56 TB de capacidad de almacenamiento administrada y el costo estimado del servicio de transmisión de Kinesis para manejar 760 GB de datos de transmisión por secuencias. Todos los créditos DBCU modelados para el procesamiento de Azure Databricks SQL sin servidor también se compraron por adelantado anualmente. Los gastos mensuales en la nube sin servidor de Azure Databricks SQL también constaron de 99,6 TB de almacenamiento de Azure Blob y una configuración modesta de Azure Databricks (6 DBU) para manejar las operaciones ETL y de transmisión. El gasto mensual de Snowflake consistió en un estado constante de 427 créditos diarios para los almacenes virtuales (18 créditos x 12 horas al día), 38 créditos diarios para manejar las alzas de escala automática, 64 créditos diarios (8 créditos x 8 horas al día) para la carga de datos en el almacén virtual y el servicio de Snowpipe, y 99,6 TB de capacidad de almacenamiento. Los costos de BigQuery incluyeron precios modelados para 800 ranuras de capacidad comprometidas, hasta 2000 ranuras de escala automática, costos de almacenamiento de nube física optimizados (datos de más de 90 días trasladados a un almacenamiento más rentable) e inserciones de transmisión.

Costos administrativos

Los costos administrativos para cada una de las soluciones se establecieron en estimaciones de buena fe basadas en opiniones de expertos y usuarios finales y análisis de casos de estudio. ESG modeló las horas semanales o únicas esperadas que un empleado dedicaría a planificar y comprar, administrar servidores o nodos de nube virtual, software EDW y administración de bases de datos, administración de ETL y consultas y administración de usuarios y procesos. Todas las soluciones basadas en la nube proporcionaron ahorros administrativos significativos en comparación con los modelos locales de ESG. La planificación y la compra de las cuatro soluciones se simplifica mucho en comparación con la planificación de implementaciones en las instalaciones, pero AWS Redshift requeriría algo de tiempo dedicado al dimensionamiento y el análisis de la solución antes de realizar la inversión inicial anual, mientras que el dimensionamiento no ideal de BigQuery, Snowflake, o el entorno sin servidor Azure Databricks SQL se puede corregir fácilmente en cualquier momento.

Una vez implementada, la solución Redshift requeriría supervisión, configuración y ajuste de recursos virtualizados (instancias o nodos), mientras que las otras soluciones se administran por completo tras bambalinas. BigQuery fue la única solución que no requirió ajuste ni configuración del software EDW y la base de datos. BigQuery, Snowflake y Azure Databricks SQL sin servidor supuestamente fueron más fáciles de administrar desde una perspectiva ETL, administración de consultas y administración de usuarios/procesos, y esto proporcionó la mayor parte de los ahorros administrativos en nuestro análisis. La transferencia simplificada de BigQuery y la capacidad de admitir IA/AA nativos e integrarlos de forma nativa con otras herramientas proporcionó ahorros administrativos adicionales. La Figura 6 compara los costos administrativos modelados de tres años para cada una de las cuatro soluciones.

Figura 6. Gasto administrativo estimado de tres años para la situación modelada de ESG



Fuente: Enterprise Strategy Group, una división de TechTarget, Inc.

Problemas que se deben considerar

Si bien los modelos de Enterprise Strategy Group se construyen de buena fe sobre suposiciones moderadas, creíbles y validadas, ningún caso modelado representará jamás cada posible entorno. Los beneficios recibidos por una organización dependen del tamaño de la organización, la naturaleza del negocio y las capacidades del producto o servicio actual que se está utilizando, junto con muchas más variables. Basar las suposiciones para los requisitos de carga de trabajo en otros casos de uso puede no dar como resultado el mismo nivel de ahorro que hemos mostrado. ESG recomienda que realice su propio análisis de EDW disponible basado en la nube a través de pruebas de concepto.

Conclusión

Producir inteligencia empresarial de calidad y conocimientos accionables lo más rentable y rápidamente posible debe ser una prioridad principal de cada organización. Los servicios de almacén de datos basados en la nube ofrecen a las organizaciones de TI ventajas importantes en agilidad y disponibilidad, a la vez que reducen en gran medida el tiempo de obtención de información. Si una organización busca crear una solución de análisis de varias nubes, emplear una moderna estrategia de nube de datos unificados, migrar su operación de almacén de datos en las instalaciones a la nube, o aumentar las operaciones existentes con servicios EDW basados en la nube para manejar nuevas oportunidades, es importante comprender las diferencias entre las ofertas y elegir la solución que proporcione a la organización la mayor agilidad, flexibilidad e interoperabilidad con otros servicios al menor costo.

Enterprise Strategy Group (ESG) comparó el costo y las capacidades de cuatro soluciones EDW líderes basadas en la nube: BigQuery, AWS Redshift, Snowflake y Azure Databricks SQL sin servidor. ESG utilizó entrevistas con usuarios finales que tenían experiencia con varias soluciones, así como estudios de casos para formar suposiciones conservadoras con el fin de modelar el TCO necesario para satisfacer las necesidades de los requisitos de EDW de una organización modelada durante un período de tres años. Nuestro modelo predijo que BigQuery podría proporcionar un TCO hasta un 54 % menor en una solución que proporciona una operación más sencilla y mayor agilidad y escalabilidad, con capacidades nativas e interoperabilidad con otros importantes servicios basados en la nube.

BigQuery está diseñada para eliminar toda la carga física y lógica de la administración, el monitoreo, el mantenimiento y la seguridad de la infraestructura EDW, lo que permite que los recursos de la organización se centren en la inteligencia en lugar de mantener el proceso de obtención de la misma. BigQuery es la única solución completamente administrada (sin nodos físicos o virtuales para administrar y mantener) con almacenamiento autooptimizado y soporte nativo para cargas de trabajo de IA/AA. BigQuery ofrece ahorros sobre las otras soluciones de EDW basadas en la nube, ya que proporciona una operación más sencilla y administración diaria, y elimina la necesidad de ajustar el tamaño de la implementación, administrar el entorno y planificar el crecimiento de los recursos virtuales. Además, BigQuery no requiere inversión inicial y puede escalar verticalmente con facilidad para satisfacer las necesidades cambiantes del negocio.

BigQuery es la única solución de nube nativa que se diseñó desde el primer día con verdaderas capacidades de transmisión y sin servidor, y ha estado ofreciendo y perfeccionando las herramientas y los procesos a lo largo de 10 años de implementaciones de clientes. Es posible que se hayan construido otros almacenes de datos en la nube mediante el aprovechamiento de los diseños y principios que se pasaron de las soluciones locales, pero están comenzando lentamente a darse cuenta de la importancia de las capacidades sin servidor y de transmisión, y comienzan a implementarlas mediante la automatización empujada.

Las soluciones de EDW actuales deben actuar como un repositorio global de información, proporcionar la agilidad para aumentar o disminuir la escala según la demanda e integrarse sin problemas con otras funciones analíticas y operativas. Pero simplemente comparar las hojas de precios públicas no es suficiente; las organizaciones también deben considerar el costo de las personas y el proceso. ESG recomienda que cada organización realice su propio análisis para comparar soluciones, y esperamos que este informe ayude a identificar las áreas que se deben considerar al comparar los servicios de EDW basados en la nube.

Con organizaciones que miran cada vez más la optimización de costos, se debe tener en cuenta que el escalamiento automático detallado de BigQuery permite a las organizaciones operar de manera predecible con una capacidad de 30 % a 40 % menos comprometida, además, la flexibilidad de elección proporcionada para las opciones de precios de almacenamiento y procesamiento permitirá a las organizaciones controlar mejor su gasto en la nube para cada proyecto o carga de trabajo. Los cambios recientes en el modelo de precios de almacenamiento y procesamiento han beneficiado a los clientes de BigQuery con los que hablamos. Como se hizo, “los cambios en el modelo de precios de BigQuery no solo nos han ahorrado dinero, sino que han aportado más consistencia a nuestra facturación, que solía fluctuar bastante según nuestro uso”. Recomendamos encarecidamente que las organizaciones consideren BigQuery en conjunto con otros servicios basados en la nube de Google Cloud para potenciar la plataforma que proporciona información empresarial de última generación en toda su organización.

TechTarget, Inc. o sus filiales. Todos los derechos reservados. TechTarget, y el logotipo de TechTarget, son marcas comerciales o marcas registradas de TechTarget, Inc. y están registradas en jurisdicciones de todo el mundo. Otros nombres y logotipos de productos y servicios, incluidos los de BrightTALK, Xtelligent y Enterprise Strategy Group, pueden ser marcas comerciales de TechTarget o sus filiales. Todas las demás marcas comerciales, logotipos y nombres de marca son propiedad de sus respectivos propietarios.

La información contenida en esta publicación se obtuvo mediante fuentes de TechTarget, por lo que se considera confiable, pero TechTarget no garantiza su confiabilidad. Esta publicación puede contener opiniones de TechTarget, las cuales están sujetas a cambios. Esta publicación puede incluir previsiones, proyecciones y otras declaraciones predictivas que representan las suposiciones y expectativas de TechTarget en vista de la información actualmente disponible. Estas previsiones se basan en las tendencias de la industria e implican variables e incertidumbres. En consecuencia, TechTarget no garantiza la precisión de pronósticos, proyecciones o declaraciones predictivas específicas contenidas en este documento.

Cualquier reproducción o redistribución de esta publicación, en su totalidad o en parte, ya sea en formato impreso, electrónico o de cualquier otro modo, a personas no autorizadas a recibirla, sin el consentimiento expreso de TechTarget, constituye una infracción de la legislación estadounidense sobre derechos de autor y estará sujeta a una acción por daños civiles y, si procede, a enjuiciamiento penal. Si tiene alguna pregunta, comuníquese con el Departamento de Relaciones con los Clientes a cr@esg-global.com.

Acerca de Enterprise Strategy Group

Enterprise Strategy Group de TechTarget proporciona inteligencia de mercado específica y práctica, investigación de la demanda, servicios de asesoramiento de analistas, orientación estratégica de GTM, validaciones de soluciones y contenidos personalizados que apoyan la compra y venta de tecnología empresarial.

 contact@esg-global.com

 www.esg-global.com