

Google BigQuery y Tableau: prácticas recomendadas

Google BigQuery y Tableau: prácticas recomendadas

Tableau y Google BigQuery permiten a las personas analizar cantidades masivas de datos y obtener respuestas rápidas gracias a una interfaz visual fácil de usar. Al usar las herramientas conjuntamente, usted puede:

- Poner el poder de Google BigQuery en las manos de usuarios corrientes para obtener un análisis rápido e interactivo.
- Analizar miles de millones de filas en cuestión de segundos con herramientas de análisis visual, sin tener que escribir ni una línea de código y sin necesidad de administrar servidores.
- Crear en pocos minutos dashboards imponentes que se conecten con sus datos de Google BigQuery y mantengan su organización actualizada.
- Compartir informes y conocimientos en la web con Tableau Server y Tableau Online, de manera que cualquier persona pueda conectarse desde cualquier dispositivo.
- Combinar la agilidad de la nube de Google BigQuery con la asombrosa velocidad de Tableau y, así, reconocer más rápidamente el valor del proyecto.

La optimización de las dos tecnologías combinadas producirá mejoras importantes en el rendimiento, ciclos de diseño más cortos y ayudará a los usuarios y las organizaciones a lograr mayores éxitos. En este informe, trataremos las técnicas para optimizar el modelado de datos y la formación de consultas a fin de maximizar la capacidad de respuesta de las visualizaciones. Además, trataremos las técnicas para obtener la mejor relación costo-beneficio cuando se utilizan Tableau y BigQuery conjuntamente.

Martin Sleeman, administrador de productos, Tableau

Marc Lobree, consultor de productos, Tableau

Vaidy Krishnan, administrador sénior de marketing de productos, Tableau

Babu Prasad Elumala, ingeniero de soluciones, Google

Seth Hollyman, administrador de programas técnicos, Google

Tino Tereshko, ingeniero de soluciones empresariales, Google

Mike Graboski, ingeniero de soluciones, Google

Contenido

Índice	3
Descripción general de la tecnología	4
Google BigQuery	4
Tableau	5
Prácticas recomendadas para el rendimiento: Tableau.....	6
Registro de rendimiento	9
Filtros de contexto	10
Agregar medidas	10
Conjuntos.....	11
Desactivar actualizaciones automáticas.....	12
Buscar advertencias	12
Prácticas recomendadas para el costo y el rendimiento: Google BigQuery	13
Desnormalizar y unir tablas previamente con JOIN	13
Realizar particiones en tablas por fechas	14
Especificar una tabla de destino si se ejecutan varias consultas similares.....	15
Conclusión.....	16
Acerca de Tableau	17
Recursos adicionales	17
Informes relacionados	17
Explore otros recursos	17

Descripción general de la tecnología

Google BigQuery

BigQuery puede procesar petabytes de datos en pocos segundos solo con SQL, sin necesidad de ajustes ni habilidades especiales. Impulsado por Dremel, la tecnología revolucionaria de Google para el análisis de conjuntos de datos masivos, BigQuery proporciona, a un costo de apenas unos centavos por gigabyte, un nivel de rendimiento por el que antes las grandes empresas debían pagar millones de dólares.

BigQuery es un almacén de datos adecuado específicamente para ejecutar consultas de SQL con conjuntos de datos masivos, estructurados y semiestructurados. Los casos de uso y los conjuntos de datos de muestra incluyen:

- Análisis ad hoc.
- Registros web.
- Registros de equipo/servidor.
- Conjuntos de datos de Internet de las cosas.
- Comportamiento de los clientes de comercio electrónico.
- Datos de aplicaciones móviles.
- Análisis del comercio minorista.
- Telemetría de juegos.
- Datos de Google Analytics Premium.
- Cualquier conjunto de datos para el cual un sistema tradicional de administración de bases de datos relacionales (RDBMS) tarde minutos (u horas) en ejecutar una consulta en lote.

BigQuery no requiere ningún tipo de operación (NoOps) ni mantenimiento. También se integra en Google Cloud Platform. A diferencia de otras soluciones de análisis en la nube, BigQuery no le exige un clúster de servidores por adelantado. BigQuery dimensiona y proporciona los clústeres de procesamiento en tiempo de ejecución.

A medida que el tamaño de sus datos aumenta, BigQuery agrega poder de procesamiento automáticamente, y a usted no le supone ningún costo adicional.

SQL heredado y SQL estándar

Google BigQuery actualizó sus API para que usen SQL estándar, además de BigQuery SQL (ahora, SQL heredado). Por otro lado, Tableau actualizó el conector de Google BigQuery a fin de admitir el uso de SQL estándar. SQL estándar proporciona a los usuarios de BigQuery algunos beneficios como las expresiones de nivel de detalle, la validación de metadatos con más rapidez y la posibilidad de seleccionar un proyecto de facturación para su conexión.

En esta guía se asume el uso de SQL estándar. Para obtener más información acerca de la migración de SQL heredado a SQL estándar, consulte [nuestra guía de ayuda en línea sobre la migración desde SQL heredado](#) en el sitio web de Google Cloud Platform.

Tableau

Tableau ayuda a las personas a ver y comprender datos. Mediante nuestros productos de software, ponemos el poder de los datos en las manos de personas corrientes. Esto permite que una amplia población de usuarios se implique en sus datos, haga preguntas, resuelva problemas y genere valor. Sobre la base de la tecnología desarrollada en la Universidad de Stanford, nuestro producto reduce la complejidad, la inflexibilidad y los costos asociados con las aplicaciones de inteligencia de negocios tradicionales. Cualquiera que se sienta cómodo usando Excel puede aprovechar Tableau Desktop para crear visualizaciones exhaustivas e interactivas y poderosos dashboards con la interfaz de usuario de arrastrar y soltar, además de compartirlos de manera segura entre organizaciones con Tableau Server o Tableau Online.

Tableau cuenta con un conector nativo y optimizado para Google BigQuery, que es compatible tanto con la conectividad de datos en tiempo real como con las extracciones en memoria. La combinación de datos de Tableau permite a los usuarios fusionar datos de BigQuery con datos de cualquiera de las 60 fuentes compatibles. Para las visualizaciones publicadas en la nube con Tableau Server o Tableau Online, se puede mantener la conectividad directa con Google BigQuery.

Prácticas recomendadas para el rendimiento: Tableau

Aproveche Tableau 10

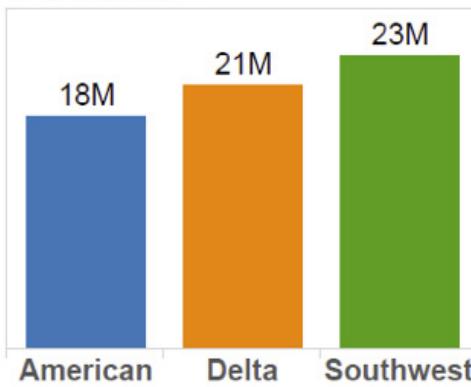
Una de las formas más sencillas de acelerar el rendimiento es asegurarse de usar Tableau 10. Si mantiene su implementación actualizada, podrá obtener todos los beneficios de rendimiento que agregamos al producto regularmente.

El lanzamiento de Tableau 9 fue un punto de inflexión para nosotros. Representó un salto gigantesco en el que introdujimos muchísimas mejoras que revolucionaron el rendimiento para garantizar que sus visualizaciones respondan adecuadamente.

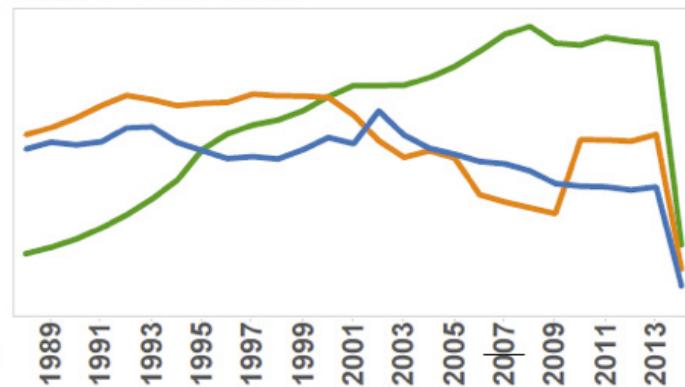
Estas mejoras incluyeron:

- **Cálculos de nivel de detalle:** Las expresiones LOD nos permiten ver más allá del nivel de detalle de la visualización. Los datos de la visualización suelen ser el resultado de filtrar la fuente. Las expresiones LOD pueden “ver” los datos antes de que estos se filtren, lo que permite hacer análisis más eficaces.
- **Consultas en paralelo:** Tableau saca ventaja de la capacidad de Google BigQuery y otras fuentes de datos de ejecutar hasta 16 consultas simultáneas. Los lotes de consultas independientes y desduplicadas se agrupan en conjunto y se envían a BigQuery si el resultado no se encuentra ya en caché. Los usuarios deberían prever grandes mejoras en el rendimiento gracias a las consultas paralelas, debido a la arquitectura flexible (escalable horizontalmente) de BigQuery.
- **Fusión de consultas:** Tableau toma varias consultas de libros de trabajo y dashboards, y las fusiona entre sí en la medida de lo posible. De ese modo, se reduce el número de consultas enviadas a BigQuery. Primero, Tableau identifica consultas similares y excluye las diferencias en las columnas devueltas. Luego, combina las consultas cuando las diferencias radican únicamente en el nivel de la agregación o en un cálculo de usuario.
- **Caché externa de consultas:** Si la fuente de datos subyacente no cambió desde la última vez que ejecutó la misma consulta, Tableau leerá automáticamente la caché de consultas previamente almacenada, lo que permite obtener tiempos de carga casi instantáneos. Por ejemplo, un libro de trabajo con una extracción de datos de Tableau de 157 millones de filas se abre 50 veces más rápido en la caché de Tableau 9 que si se abre en Tableau 8.3 sin caché.

Top Airlines



Flights Over Time



State Distribution

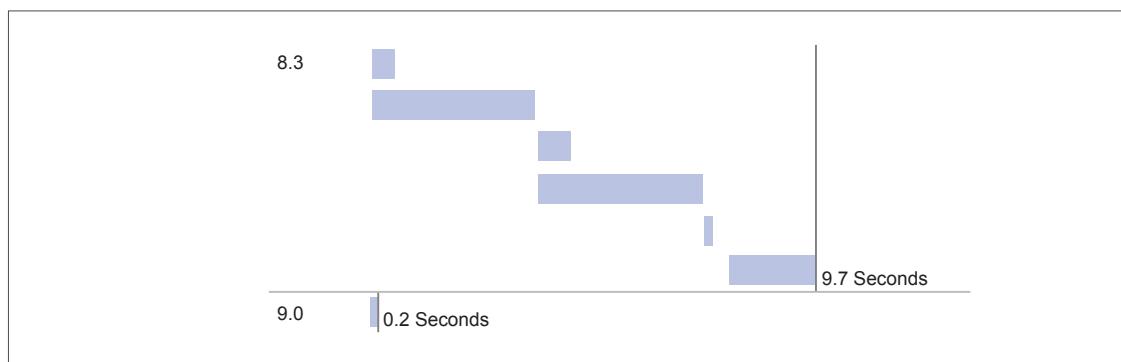
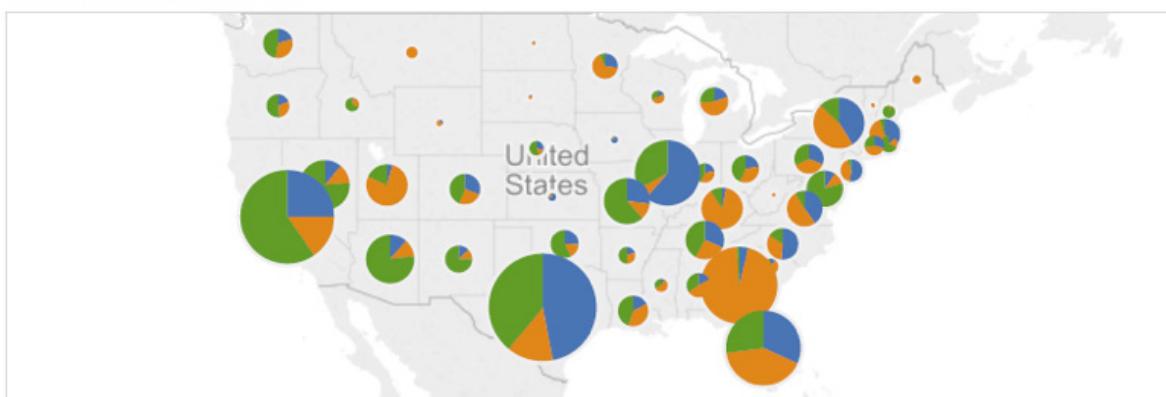


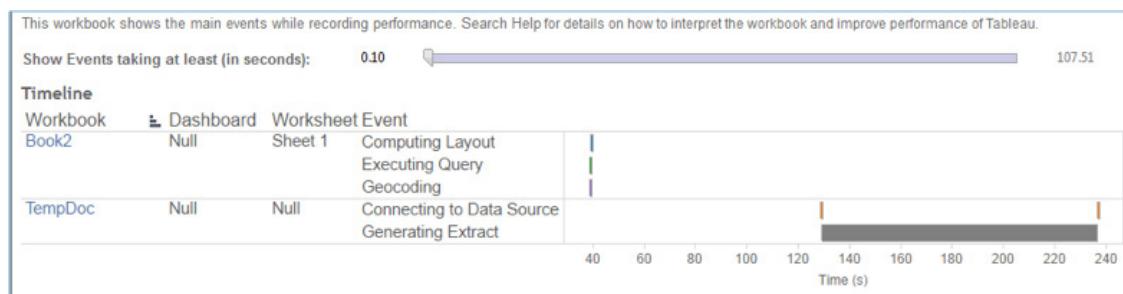
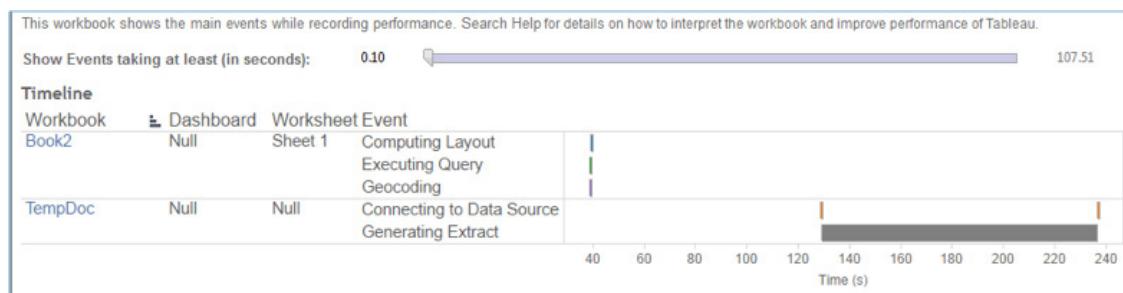
Figura 1: Mejoras de rendimiento para un conjunto de datos de 157 millones de filas abierto en Tableau 9.0 en comparación con Tableau 8.3.

En Tableau 10, continuamos innovando para garantizar el buen rendimiento de las implementaciones empresariales escalables. En particular, incorporamos:

- **Rendimiento de navegador mejorado:** En Tableau 10, guardamos en caché el libro de trabajo que se carga inicialmente para que este responda con mayor velocidad. Disfrute de tiempos de carga más rápidos, actualizaciones instantáneas basadas en sus interacciones y mucho más.
- **Conexiones a pedido en Tableau Desktop:** Cuando abre un libro de trabajo publicado, Tableau Desktop solo se conecta con las fuentes de datos necesarias para visualizar los datos de la hoja actual. En otras palabras, puede ver los datos de manera mucho más rápida.
- **Mejoras relacionadas con la estabilidad de Tableau Server:** Tableau 10 incluye muchas mejoras relacionadas con la estabilidad de Tableau Server. Optimizamos Tableau Server para la resolución de problemas durante los períodos de alta latencia de E/S de disco en las instalaciones de servidor único. Tableau Server requiere tres nodos para ofrecer alta disponibilidad. Sin embargo, en el pasado permitimos la configuración de la comutación tras error y la replicación con dos nodos, lo que confundía a los clientes. Las instalaciones de dos nodos se limitan a una única instancia del repositorio. Si necesita comutación tras error o alta disponibilidad en una segunda instancia del repositorio, instale Tableau Server en un clúster de al menos tres nodos. De ese modo, puede configurar dos instancias del repositorio y obtener los beneficios de una comutación automática tras error. También redujimos el uso general que los procesos de Tableau Server hacen de la memoria para mejorar el rendimiento.

Registro de rendimiento

El registro de rendimiento es una poderosa herramienta incorporada que le permite identificar consultas lentas y optimizar sus libros de trabajo para obtener el máximo rendimiento. Esto se realiza mediante el registro de tiempo transcurrido entre que un libro de trabajo individual ejecuta una consulta y calcula el trazado. Al pasar el puntero del mouse sobre una de las barras verdes de más abajo, se mostrará al usuario la consulta que se está generando en BigQuery. Luego de identificar una consulta lenta, a menudo puede resolver el problema de rendimiento revisando su modelo de datos.



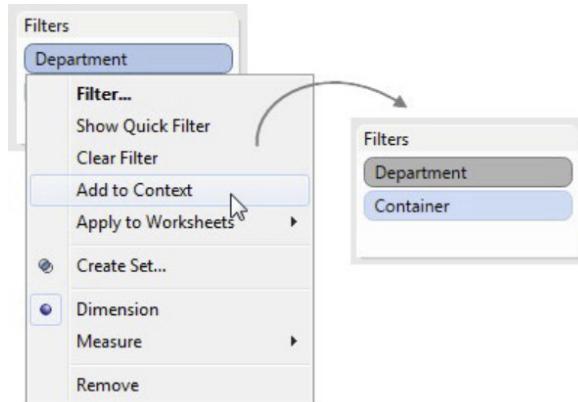
Para obtener instrucciones sobre cómo crear o interpretar un registro de rendimiento, haga clic en uno de los enlaces siguientes:

- [Registrar y analizar el rendimiento de un libro de trabajo](#)
- [Interpretar un registro de rendimiento](#)

Filtros de contexto

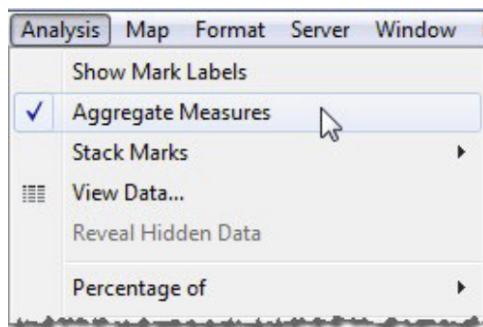
Si está aplicando filtros a una fuente de datos grande, puede configurar filtros de contexto para mejorar el rendimiento. En primer lugar, se aplica un filtro de contexto a la fuente de datos, de manera que los filtros adicionales se apliquen solo sobre los registros resultantes. Esta secuencia evita tener que aplicar un filtro a cada registro en la fuente de datos.

Si está configurando filtros que reducen considerablemente el tamaño del conjunto de datos y piensa usarlos en muchas vistas de datos, debe configurar esos filtros como filtros de contexto. [Consulte nuestra guía de ayuda en línea sobre cómo mejorar el rendimiento de las vistas con filtros de contexto](#) para saber cómo crear filtros de contexto.



Agregar medidas

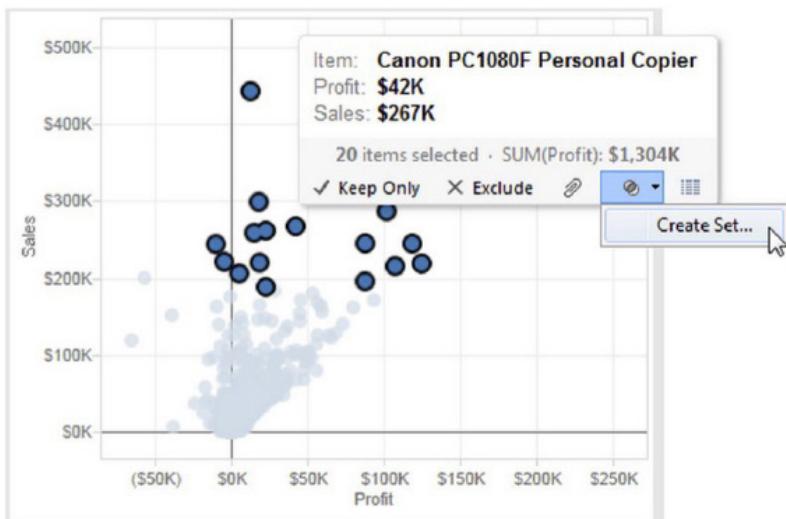
Si las vistas que genera son lentas, asegúrese de estar trabajando con medidas agregadas en vez de medidas desagregadas. Cuando las vistas son lentas, por lo general, esto quiere decir que está intentando ver demasiadas filas de datos al mismo tiempo. Puede disminuir la cantidad de filas mediante la agregación de datos. En otras palabras, asegúrese de que la opción Agregar medidas en el menú Análisis esté seleccionada. Para obtener más información, consulte [nuestra guía de ayuda en línea sobre la agregación de datos](#).



Conjuntos

Si desea filtrar una dimensión para quitar miembros sobre la base de un rango de valores de medidas, debe crear un conjunto, en vez de utilizar un filtro cuantitativo. Por ejemplo, puede crear un conjunto que solo devuelva los 50 primeros elementos de una dimensión, en lugar de todos sus elementos. Para obtener más información, consulte [nuestra guía de ayuda en línea sobre la creación de conjuntos](#).

Cuando cree un grupo desde una selección, según se describe en [nuestra guía de ayuda en línea sobre la creación de grupos](#), asegúrese de incluir solo las columnas que le interesen. Cada columna adicional incluida en el conjunto tendrá como resultado una reducción en el rendimiento.



Agregar filtros primero

Si está trabajando con una fuente de datos grande y tiene las actualizaciones automáticas desactivadas, es posible generar una consulta muy lenta cuando se agregan filtros a la vista. En vez de crear la vista y, luego, especificar los filtros, primero debe especificar los filtros y, después, arrastrar los campos a la vista. De esa manera, cuando ejecute la actualización o active las actualizaciones automáticas, se evaluarán primero los filtros.

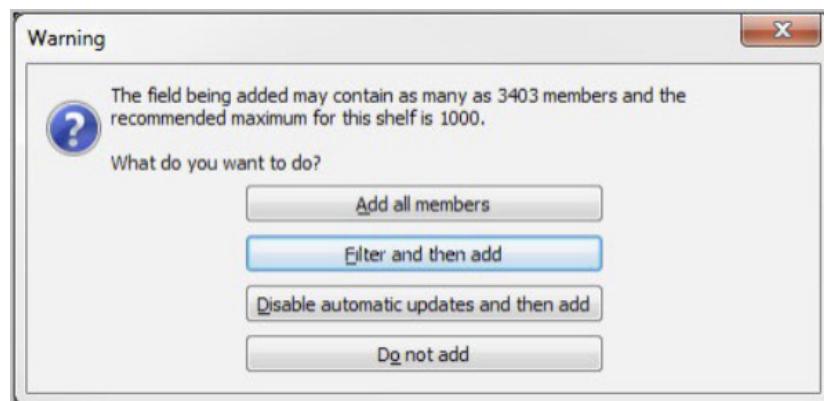
Desactivar actualizaciones automáticas

Cuando coloca un campo en un estante, Tableau genera la vista mediante consultas automáticas a la fuente de datos. Si está creando una vista de datos densa, las consultas pueden llevar mucho tiempo y degradar considerablemente el rendimiento del sistema. En este caso, puede dar la instrucción a Tableau de desactivar las consultas mientras crea la vista. Luego, puede volver a activar las consultas una vez que haya finalizado para ver los resultados. Para obtener más información, consulte [nuestra guía de ayuda en línea sobre la administración de consultas](#).

Buscar advertencias

Tableau muestra un cuadro de diálogo de advertencia sobre el rendimiento cuando intenta colocar una dimensión grande (con muchos miembros) en algún estante. El cuadro de diálogo proporciona cuatro opciones, según se muestra en la figura a continuación.

Si elige agregar todos los miembros, es posible que observe una degradación importante del rendimiento.



Prácticas recomendadas para el costo y el rendimiento: Google BigQuery

Para hacer consultas de alto rendimiento y reducir costos, se recomienda evitar el uso de tablas federadas, que toman datos de una fuente externa, como Google Cloud Storage. En esos casos, si desea realizar consultas iterativas en el conjunto de datos, debe usar la API de consulta para materializar los datos en BigQuery (independientemente de Tableau) a fin de posibilitar las consultas de alto rendimiento sobre el conjunto de Tableau.

Desnormalizar y unir tablas previamente con JOIN

BigQuery puede ejecutar uniones muy grandes con la función JOIN, manteniendo un excelente rendimiento de dicha función. BigQuery es un almacén de datos en columnas, y su rendimiento máximo se logra trabajando con conjuntos de datos desnormalizados.

Puesto que el almacenamiento de BigQuery es muy poco costoso y muy escalable, a menudo, es prudente desnormalizar y unir previamente con JOIN los conjuntos de datos en tablas homogéneas. Básicamente, está intercambiando recursos de cálculo por recursos de almacenamiento (estos últimos más rentables y de mayor rendimiento). BigQuery es un almacén en columnas. Por lo tanto, prescindir de recursos de cálculo para tener más capacidad de almacenamiento es conveniente, ya que facilita la compresión de los datos.

BigQuery es una herramienta ETL excelente, que le permite ejecutar canalizaciones y transformaciones masivas de manera rápida y efectiva. Asegúrese de habilitar la opción “Allow Large Results” (Permitir resultados grandes) cuando materialice conjuntos de datos de más de 128 MB.

Para obtener más información sobre cómo preparar los datos para su carga y cómo hacer consultas sobre estos con [el lenguaje SQL de BigQuery](#), consulte los documentos en línea que aparecen a continuación.

- cloud.google.com/bigquery/preparing-data-for-bigquery#denormalizingdata
- cloud.google.com/bigquery/querying-data#largequeryresults

Realizar particiones en tablas por fechas

Por su naturaleza, algunos datos pueden particionarse por fecha. Por ejemplo, los datos de registros o cualquier dato cuyos registros incluyan una marca de hora con un incremento repetitivo. En este caso, realice particiones de sus tablas de BigQuery por fecha e incluya la fecha en el nombre de la tabla. Si desea sacar ventaja de esto, debe aprovechar SQL personalizado en Tableau. Para obtener más información, consulte [nuestra guía de ayuda en línea sobre la conexión a una consulta de SQL personalizada](#). Por ejemplo, ponga a su tabla un nombre del siguiente tipo:

`mytable_20170501, mytable_20170502, etc.`

Luego, cuando quiera ejecutar una consulta que filtre por fecha, use la función de comodín para tablas de BigQuery:

```
SELECT
    nombre
FROM
    `myProject.myDataSet.mytable_*`
WHERE
    edad >= 35
```

El ejemplo anterior incluirá automáticamente todas las tablas con el prefijo mytable_.

Para usar comodines, sus tablas deben tener nombres que cumplan con este patrón:

[prefijo arbitrario]AAAAMMDD.

Otros sistemas de bases de datos dependen del particionamiento para mejorar el rendimiento. Realizar particiones por fecha, en realidad, genera una diferencia insignificante en el rendimiento de BigQuery. El factor principal aquí, sin embargo, es el costo. Como está procesando menos datos, está pagando menos dinero por consulta.

Advertencia: Si decide particionar en niveles diminutos, es posible que obtenga demasiadas partes, lo que afectaría directamente al rendimiento. Debe tener cuidado de no particionar demasiado a la vez. En general, cualquier partición superior a una división diaria es aceptable.

Para comprender mejor cómo trabajar con comodines, haga clic [aquí](#).

Especificar una tabla de destino si se ejecutan varias consultas similares

Mientras que el almacenamiento de consultas en caché es útil si está ejecutando muchas consultas idénticas, no será de ayuda si está ejecutando consultas similares, pero ligeramente diferentes (p. ej., cambiar solo los valores en una cláusula WHERE entre las ejecuciones de consultas). En este caso, ejecute una consulta en su tabla fuente y guarde los registros que consultará reiteradamente en una nueva tabla de destino. Luego, ejecute las consultas en la nueva tabla de destino que creó.

Por ejemplo, digamos que está planeando ejecutar tres consultas con tres cláusulas WHERE diferentes:

WHERE col1 = "a"

WHERE col1 = "b"

WHERE col1 = "c"

Ejecute una consulta en su tabla fuente y guarde los registros resultantes en una tabla de destino:

```
SELECT col1
```

```
FROM source
```

```
WHERE col1 = "a" OR col1 = "b" OR col1 = "c"
```

Al vincular las cláusulas WHERE con el operador “OR”, se capturan todos los registros pertinentes. Nuestra nueva tabla de destino es potencialmente mucho más pequeña que la tabla fuente original. Puesto que el precio de BigQuery se basa en la cantidad de datos procesados por consulta, la ejecución de consultas sucesivas en la nueva tabla de destino permite ahorrar dinero, al contrario de lo que sucede al ejecutar las consultas directamente en la tabla fuente. En el futuro, procure limpiar estas tablas para evitar el aumento del costo de almacenamiento.

Conclusión

Al aplicar las prácticas recomendadas, tanto los usuarios corporativos como los analistas de datos podrán maximizar el rendimiento y la capacidad de respuesta de las visualizaciones de Tableau generadas a partir de Google BigQuery. Cuando estas tecnologías se combinan, los usuarios pueden realmente visualizar miles de millones de filas de datos a la velocidad del pensamiento.

Acerca de Tableau

Tableau ayuda a las personas a ver y comprender los datos. Tableau ayuda a todas las personas a analizar, visualizar y compartir información rápidamente. Más de 29 000 cuentas de clientes obtienen resultados rápidos con Tableau en la oficina o en cualquier otro lugar. Además, decenas de miles de personas usan Tableau Public para compartir datos en sus blogs y sitios web. Para ver la forma en que Tableau puede ayudarlo, descargue la versión de prueba gratuita en tableau.com/es-es/trial.

Recursos adicionales

[Descargar versión de prueba gratuita](#)

Informes relacionados

[¿Por qué análisis de negocios en la nube?](#)

[Las 10 tendencias principales de la nube para 2017](#)

[Tableau Server y Google Cloud Platform: inteligencia de negocios ultrarrápida en la nube \(en inglés\)](#)

[Ver todos los informes](#)

Explore otros recursos

[Demostraciones de productos](#)

[Capacitación y tutoriales](#)

[Comunidad y soporte](#)

[Historias de clientes](#)

[Soluciones](#)

Tableau y Tableau Software son marcas comerciales de Tableau Software, Inc. Todos los demás nombres de productos y empresas pueden ser marcas comerciales de las respectivas empresas con que se asocian.

