



# *Reporting* Rentable con Bases de Datos Replicadas

Libro Blanco Dbvisit

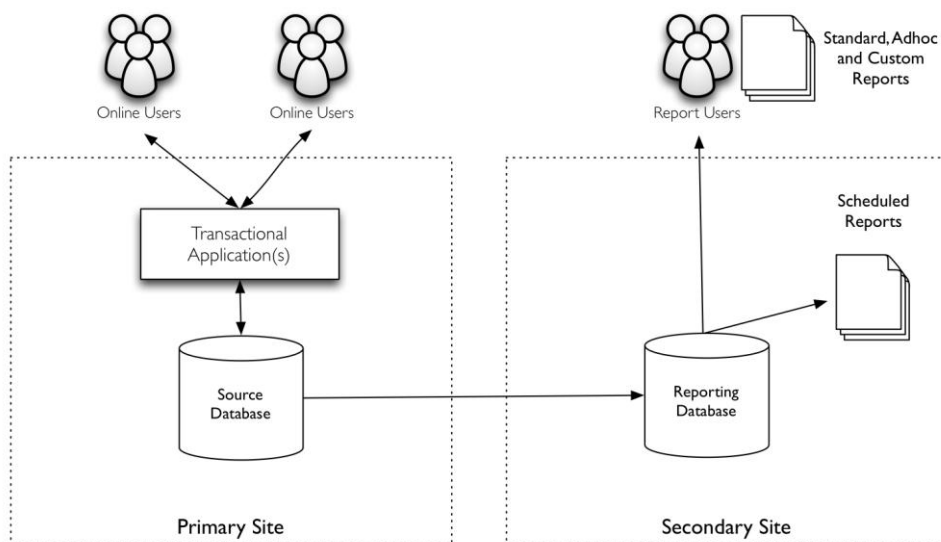
## Resumen

Hoy día las empresas dependen más que nunca del acceso en tiempo real a información vital producida por sus sistemas en línea. En cualquier compañía esta información proviene de diversas fuentes, y se crea y actualiza constantemente mediante una amplia gama de sistemas. Las compañías exigen poder informar y revisar al detalle estos datos, prácticamente a medida que se crean.

Los sistemas en línea que crean y mantienen estos datos deben estar disponibles las 24 horas, siete días a la semana; no sólo para los usuarios corporativos, sino también para socios de negocio y clientes. Para asegurarse de que los sistemas funcionan en caso de desastre utilice herramientas como Oracle Data Guard para Oracle Enterprise Edition y Dbvisit Standby para Oracle Standard Edition. Estos productos garantizan la protección de bases de datos en caso de desastre. Pero los departamentos de TI quieren seguir respondiendo con rapidez y funcionando profesionalmente, incluso ante la creciente demanda de *reporting*.

La obtención de informes puede ser una actividad intensiva de hardware,, por lo que simplemente apuntar con un generador de informes a una base de datos transaccional en línea para la elaboración de informes es una receta para el desastre. Si el sistema en línea no se detiene por completo, se ralentizará por la actividad del sistema de *reporting*. Al mismo tiempo, la capacidad de crear y recrear informes se verá rápidamente afectada por la estructura y puesta a punto de la base de datos en línea, así como por la pugna con los sistemas en línea por los recursos informáticos.

Una forma de equilibrar estas demandas, aparentemente en conflicto, es instalar una versión separada de la base de datos de producción utilizada específicamente para crear informes. El siguiente diagrama ilustra el concepto de una base de datos de *reporting* separada que se mantiene actualizada con los datos de la base transaccional de los sistemas de producción.



**Figure 1.** Implementación de base de datos de informes

## Libro blanco Dbvisit

En este entorno, la base de datos de *reporting* contiene una réplica actualizada de los datos de las bases de producción mantenidas a través del sistema transaccional en línea. Esta base de datos de *reporting* se utiliza para crear y generar informes en línea, cuadros de mando visuales y ejecutar aplicaciones de inteligencia empresarial. También se puede utilizar para alimentar *data warehouses* y dar soporte a prácticamente cualquier otra forma de acceso de sólo lectura, incluyendo búsquedas en línea.

Al habilitar la extracción de información en tiempo real, en paralelo con el funcionamiento continuo de los sistemas transaccionales en línea, se hace posible mantener el rendimiento de estos sistemas al mismo tiempo que se crean informes sobre datos en tiempo real de una manera eficiente.

Las técnicas de replicación de datos posibilitan la creación de un entorno en el que se identifiquen cambios en los datos del entorno transaccional a medida que ocurren, y se apliquen a una base de datos de informes en hardware independiente. Es entonces cuando estos datos pueden diseñarse y ajustarse para dar soporte a la actividad de *reporting* en concreto.

Dbvisit Replicate es una herramienta diseñada específicamente para obtener bases de datos de *reporting* actualizadas sin afectar a la base de datos origen. Dbvisit Replicate utiliza la replicación de datos para sincronizar información a través de bases de datos, permitiendo que las actividades de *reporting* se realicen en una plataforma alternativa.

Dbvisit Replicate utiliza la tecnología de *mining*, *change data capture*<sup>1</sup> (CDC) altamente eficiente, para detectar cambios en los datos de origen y replicar y distribuir estos cambios a través de las bases de datos Oracle y no-Oracle. El proceso de exploración o *mining* se ejecuta fuera de la base de datos, lo que significa que no es intrusivo y no afecta al rendimiento del sistema en línea.

<sup>1</sup> Tenga en cuenta que esto se refiere al propio mecanismo *Change Data Capture* de Dbvisit y no al producto de Oracle llamado "Oracle CDC".

## ¿Por qué crear una base de datos de *reporting*?

Para las empresas es cada vez más importante proporcionar acceso las veinticuatro horas del día a sus sistemas en línea. Tanto los usuarios corporativos como proveedores, socios comerciales y clientes exigen acceso continuo y en línea a los sistemas. Al mismo tiempo, las compañías tienen que ser capaces de crear informes en tiempo real, basados en los datos que esos sistemas contienen.

Para lograr esto sin afectar al rendimiento y la fiabilidad de los sistemas en línea, es posible crear una copia de la base de datos de producción (fuente) para uso exclusivo de los redactores de informes y otras funciones de "sólo lectura".

La creación de una base de datos de *reporting* independiente, que se utilice exclusivamente para la presentación de informes, ofrece una serie de ventajas:

- 1) El *Reporting* no afecta al rendimiento de la base de datos de producción, por lo que las aplicaciones en línea continúan funcionando con la misma capacidad de respuesta y rendimiento al que los usuarios están acostumbrados.
- 2) El *Reporting* no impacta en la fiabilidad de la base de datos de producción.
- 3) La base de datos en línea puede ajustarse específicamente para el alto volumen que el acceso transaccional requiere para dar soporte a las aplicaciones y sus usuarios.
- 4) La base de datos de *reporting* puede adaptarse específicamente para cumplir con las actividades de creación de informes.
- 5) El contenido y la estructura de la base de datos de *reporting* pueden adaptarse a las necesidades del sistema de información, por lo que es más fácil para el usuario crear y mantener informes, además esto aumenta la eficiencia del proceso de creación de informes.
- 6) La base de datos de *reporting* puede instalarse en sistemas operativos y hardware no propietarios.
- 7) La descarga de la actividad de *reporting* a una plataforma separada de hardware dedicada, puede eliminar la necesidad de costosas actualizaciones de hardware para los sistemas en línea. Esto puede llevarse a cabo incluso en un entorno basado en *cloud*, como el servicio ofrecido por AWS de Amazon.
- 8) La base de datos de *reporting* puede instalarse en una tecnología de base de datos diferente a la base de datos fuente, lo que proporciona flexibilidad en la elección de las herramientas de informes. Esto permite a las empresas aprovechar las herramientas de redacción de informes y el conocimiento existentes.
- 9) Si se desea, pueden suministrarse múltiples copias de bases de datos de *reporting* en diferentes lugares.
- 10) Las bases de datos de *reporting* no siempre necesitan una copia completa de todas las tablas de la base de datos de producción. Se pueden filtrar esquemas, tablas e incluso filas dentro de tablas.

## Libro blanco Dbvisit

Puede referirse a una base de datos de *reporting* con distintos términos, entre los que se incluyen:

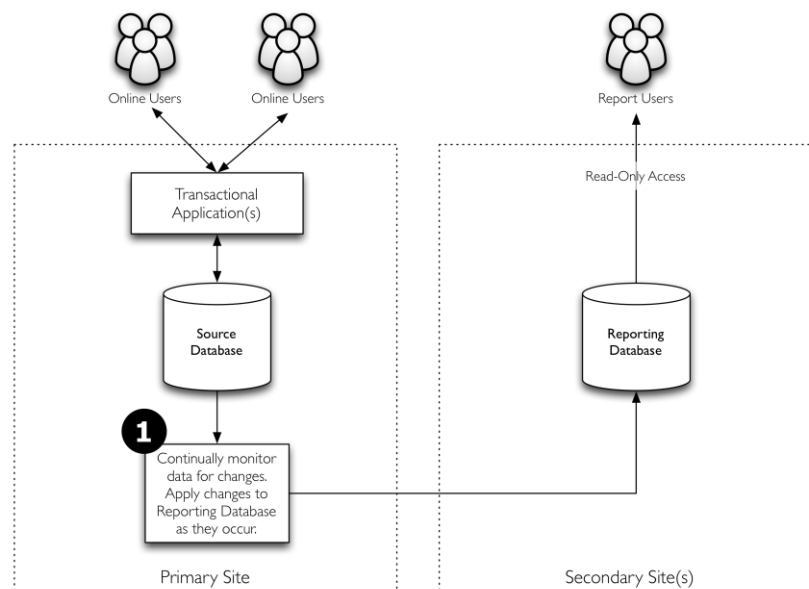
- *Offload reporting database*
- Base de datos *ad-hoc*
- *Reporting Operativo*
- Réplicas de lectura
- *Query offloading*
- Servicios de bases de datos de *Reporting*

Sin embargo, todos tienen el mismo propósito: proporcionar una copia secundaria de la base de datos transaccional para su uso en la visualización y análisis de información en la base de datos.

## Replicación de datos

Necesitamos una solución que permita la actualización en línea para sistemas transaccionales al mismo tiempo que la extracción de datos de la base de *reporting*, y que además pueda hacer esto de manera continua. Ni la actualización completa de datos ni un *ETL* constituyen una opción viable; ambos enfoques causan retrasos en la aplicación de datos actualizados en la base de datos de *reporting* y ambos afectan a la disponibilidad y el rendimiento del sistema de origen.

El tercer enfoque que aquí se presenta, la replicación de datos, permite el seguimiento continuo de los cambios en los datos de origen a través de los registros de transacciones que mantiene la base de datos Oracle. A medida que el proceso de replicación identifica los cambios en estos registros también extrae los detalles de los cambios y los aplica a la base de datos de *reporting* de destino como sentencias SQL.



**Figure 2.** Proceso de replicación de datos

## Libro blanco Dbvisit

La propuesta de la replicación ofrece una serie de ventajas significativas sobre los enfoques alternativos:

1. Carga muy baja en el entorno de la base de datos y servidor fuente
2. Replicación en la base de *reporting* de destino casi en tiempo real de los cambios en los datos
3. Acceso ininterrumpido al sistema fuente para usuarios en línea y otra actividades transaccionales
4. Sin restricciones para cambiar los datos fuente o el sistema transaccional en línea
5. Los datos borrados se detectan fácilmente y se replican a los datos destino
6. Los datos fuente y de destino pueden tener diferentes configuraciones de índices para una operación óptima en función de cada necesidad

El acceso en línea a sistemas transaccionales no se ve interrumpido para los usuarios corporativos, los trabajadores móviles, socios y clientes, ni siquiera mientras el proceso de replicación de datos se ejecuta sobre los datos origen. La siguiente ilustración muestra la disponibilidad constante de las aplicaciones en línea, junto con la replicación continua de datos (con lo que se logra que no haya retrasos en la actualización de datos):

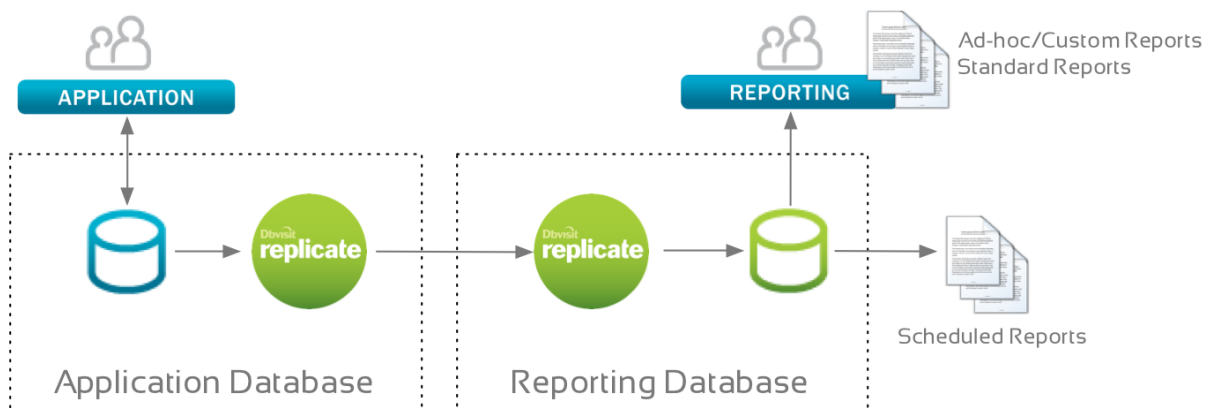
## Estructura de Dbvisit Replicate

### Resumen

Dbvisit Replicate utiliza la replicación de datos para sincronizar información a través de bases de datos, permitiendo que las actividades de *reporting* se descarguen en una plataforma alternativa. Esto proporciona un beneficio significativo a nivel de infraestructura, facilitando el *reporting* de negocio en tiempo real y la reducción del impacto del *reporting* en sistemas en línea.

Dbvisit Replicate es un software potente y asequible, fácil de instalar y configurar. Está diseñado específicamente para la creación y actualización de bases de datos de *reporting*. Dbvisit Replicate ofrece procesos de replicación de Oracle a Oracle, MSSQL y MySQL y viene con todas las herramientas y servicios necesarios para lograr un uso fácil, incluyendo la detección y resolución de conflictos.

Dbvisit Replicate utiliza una tecnología de exploración *change data capture (CDC)*, altamente eficiente, para detectar cambios en los datos de origen y replicar y distribuir estos cambios a través de las bases de datos Oracle así como en aquellas que no lo sean. El proceso de exploración o *mining* se ejecuta fuera de la base de datos, lo que significa que no es intrusivo y no afecta al rendimiento del sistema en línea.<sup>3</sup>



**Figure 3.** Información general sobre Dbvisit Replicate

La replicación de datos mediante tecnología *CDC* y basada en los archivos de Bases de datos *redo log* ofrece muchas ventajas, entre ellas:

- Los datos para el *Reporting* están siempre actualizados y disponibles

<sup>3</sup> Tenga en cuenta que esto se refiere al propio mecanismo *Change Data Capture* de Dbvisit y no al producto de Oracle llamado "Oracle *CDC*".

## Libro blanco Dbvisit

- Impacto mínimo en la base de datos fuente
- Sin necesidad de cambios de programas en las aplicaciones que utilizan la base de datos
- Baja latencia en los cambios de la base de datos fuente
- La integridad transaccional se mantiene mediante la reproducción de las transacciones originales en el orden en que se realizaron
- No se requieren cambios en el esquema de la base de datos (por ejemplo, no hay *triggers* ni *materialized views* etc.).

Las principales características de Dbvisit Replicate incluyen:

- Proceso de instalación guiado paso a paso por medio de un asistente
- Replicación en tiempo real de bases de datos Oracle a bases de datos de destino Oracle, MSSQL y MySQL
- Los datos a replicar pueden definirse en la base de datos, a nivel de esquema o de tabla
- Monitorización en tiempo real del estado del proceso
- Notificación mediante correo electrónico
- Intercambio seguro de datos
- Funcionamiento sin la intervención del operador

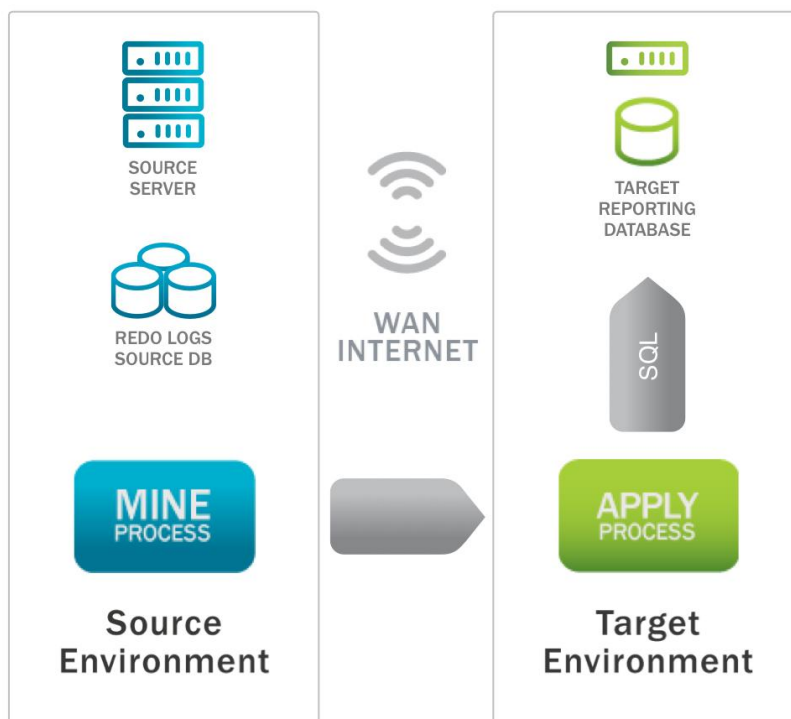
## ¿Cómo funciona?

Dbvisit Replicate permite la replicación de datos en tiempo real entre Oracle y otras bases de datos distintas. Dbvisit Replicate utiliza tecnología *change data capture* basada en *redo log* para detectar cambios en la base de datos origen y replicar y distribuir, en tiempo real, los datos modificados a Oracle y a otras bases de datos. Esto también se conoce como *redo log mining*. Dbvisit Replicate utiliza su propia tecnología *redo log mining*, no dependiendo de ningún software de terceros.

Además de los procesos integrados de detección y resolución de conflictos, el almacenamiento en caché de Dbvisit Replicate asegura que todos los registros *committed* actualizan de manera segura la(s) base(s) de datos destino, lo que garantiza que la base de datos *de reporting* permanece actualizada y disponible incluso en caso de un corte de luz.

Dbvisit Replicate lee (*mines*) el *redo log* de Oracle en tiempo real en la base de datos fuente, los convierte en comandos SQL y los ejecuta (*applies*) en la base de datos destino.





**Figure 4.** Estructura de Dbvisit Replicate

Dbvisit Replicate es muy eficiente, no sobrecarga el sistema y puede gestionar grandes volúmenes de transacciones sin comprometer el rendimiento de la base de datos fuente.

- El proceso **Mine** lee los Oracle *redo log* en la base de datos fuente en tiempo real y convierte la información de estos registros en un formato interno para su posterior aplicación a la base de datos de *reporting*.
- El proceso **Apply** toma el fichero de datos interno creado por el proceso *Mine* y lo convierte en SQL, para luego ejecutarlo en la base de datos de informes destino.

## Subconjuntos de datos, partición y filtrado

Dbvisit Replicate ofrece varias funciones que se pueden utilizar para controlar la estructura y el contenido de la base de datos de *reporting* destino. Esto permite la configuración de una solución que maximice la eficiencia y el rendimiento, a la vez que garantiza que todos los datos necesarios estén disponibles para suministrar la información necesaria.

En muchas ocasiones, solo tiene que replicarse un subconjunto o esquema de la base de datos fuente en la base de destino para permitir el *reporting*. De hecho, por razones de rendimiento, y para simplificar la creación y mantenimiento de informes, es mejor que solo se transfieran los datos necesarios para los

## Libro blanco Dbvisit

informes. Dbvisit Replicate permite esta opción al dejar que un administrador restrinja los datos que se replican.

Mediante la exclusión del proceso de replicación de tablas completas es posible restringir los datos a nivel de esquema. Dbvisit Replicate excluye automáticamente las tablas temporales, mantenidas para las aplicaciones fuente, como por ejemplo aquellas que son ampliamente utilizadas por los sistemas ERP.

La funcionalidad *Data partitioning* permite la restricción de información a nivel de tabla:

- *Vertical partitioning* permite la exclusión de columnas específicas del proceso de replicación, lo que puede ser particularmente útil para tablas grandes o que no sean necesarias para el *Reporting* en la base de datos de destino. Por ejemplo, es posible que queramos excluir datos BLOB, así como valores internos específicos de la aplicación (como *internal keys*).
- *Horizontal partitioning* permite el filtrado y replicación de solo aquellas filas dentro de la base de datos fuente que coincidan con los criterios específicos. Por ejemplo, replicar solo datos en los que *región* = "Noroeste". Esto permite la configuración de una base de datos de *reporting* en el lugar de destino especialmente confeccionada para dar soporte a un conjunto específico de informes. También se puede utilizar para crear múltiples bases de datos de informes en destino basadas en contenido, cada una con un conjunto diferente de datos.

A través de una combinación de filtrado de tablas y de *horizontal and vertical partitioning*, es posible gestionar un proceso de replicación altamente perfeccionado que suministre los datos requeridos a la base de datos de *reporting* de forma eficiente y rentable.

El *mapping* de datos amplía las capacidades de réplica aún más al permitir que de la información de la base de datos fuente se haga *mapping* en una ubicación alternativa en la base de datos de destino. El *mapping* permite el control de los datos que se van a replicar de la base de datos fuente, y también sobre la estructura de los datos aplicados a la base de datos de destino. Puede modificarse el orden de las columnas dentro de una tabla.

## *Optimistic Commits*

Dbvisit Replicate utiliza el mismo principio de *optimistic commits* que aplica Oracle. Este enfoque asume que la mayoría de las transacciones serán *committed*, y comienza a escribir las transacciones en el disco antes de que el verdadero *commit* se haya ejecutado. Oracle utiliza este mismo método de *optimistic commit* en la propia base de datos. En Dbvisit Replicate esto significa que con los datos se hace el *mining* y el *applied* **antes** del *commit*.

Para las transacciones grandes esto puede resultar beneficioso, ya que las transacciones no se almacenan sino que se hace *mine* y *apply* de ellas tan pronto como se escriben en los *redo logs* de la base de datos fuente de Oracle. Esto también puede reducir la cantidad necesaria de disco y memoria mientras se espera por un *commit*, así como reducir la brecha entre los procesos de *mining* y *apply* a unos pocos segundos.

## Libro blanco Dbvisit

Obviamente, mientras las transacciones se escriben en la base de datos de *reporting*, se aplican a la base de datos fuente, los cambios no son visibles en la base de datos de *reporting* hasta que se haya aplicado el *commit* en la base de datos fuente y replicado en la base de datos de *reporting* (transacciones autónomas).

### Configuración inicial - *Data Instantiation*

Antes de empezar la replicación, la base de datos destino debe tener una copia de los objetos que vayan a ser replicados (esto se puede denominar "instantiating" o "seeding" en la base de datos destino). Es altamente deseable que los datos en los objetos que se vayan a replicar estén también completamente *in sync*; si los datos no están *in sync* antes de iniciar la replicación, existe una alta posibilidad de que se produzcan conflictos de datos (por ejemplo, si un registro se actualiza o se elimina de la base de datos fuente, pero no existe en la base de datos de destino).

Dbvisit Replicate puede ayudar, creando y estableciendo puntos de arranque para asegurar que los cambios no se pierden una vez empiece la replicación. Dbvisit Replicate ofrece algunos métodos integrados diferentes como Oracle Data Pump, o, para versiones más antiguas de Oracle, utiliza *imp* y *exp*. No obstante, cualquier herramienta que pueda crear una carga de datos, como el *SCN*, servirá.

Es también fundamental que la replicación comience en el punto en el que las bases de datos fuente y destino estén *en sync*, asegurando que no se pierden cambios cuando se inicia la replicación. En bases de datos Oracle, Dbvisit Replicate gestiona esta *instantiation* de forma automática, utilizando *Oracle System Change Number (SCN)* para determinar el punto en el que la tabla está "preparada". Para bases de datos de destino no Oracle, esta *instantiation* debe coordinarse de forma manual como parte de la instalación para la base de datos de *reporting*.

### Detección y resolución de conflictos

Los conflictos y la resolución de conflictos son importantes en la replicación de bases de datos. Cuando se produce un conflicto significa que existe una posibilidad de que los datos entre las bases de datos fuente y destino no estén sincronizados. Esto también se conoce como divergencia de datos.

Dbvisit Replicate ofrece funciones configurables para la detección y resolución de conflictos:

- La detección de conflictos nos garantiza que cuando exista la posibilidad de divergencia de datos se active una notificación o alerta.
- La resolución de conflictos nos asegura que cuando se produzca un conflicto se suministre suficiente información para que se pueda resolver.
- La gestión de conflictos establece reglas predefinidas para establecer pautas para cuando se produzca un conflicto.
- La capacidad para el administrador de crear una lista de conflictos, que permite ver el SQL que está causando el conflicto.

## Libro blanco Dbvisit

Todos los procesos de detección, gestión, resolución y listado de conflictos se realizan desde la consola centralizada de Dbvisit Replicate.

### La detección de un conflicto

Para detectar que la información en destino ha sido modificada independientemente de los datos fuente, Dbvisit Replicate incluye una columna de cambios o actualizaciones en el apartado *WHERE*. Esto permite al sistema detectar inconsistencias entre las etapas *mine* y *apply*. Cualquier discrepancia entre el estado actual de la fila y el estado esperado indica divergencia de datos. También garantiza la notificación al usuario y que no se pierda ningún dato del histórico.

### Gestión de un conflicto

Como parte de la configuración, Dbvisit Replicate está parametrizado para controlar su respuesta a los diferentes tipos de conflictos mediante una amplia gama de gestores de conflicto integrados. Se puede especificar un enfoque de resolución de conflictos independiente para cada tabla replicada y para cada tipo de operación, incluyendo la gestión del caso especial en que cero filas, o más de una fila, se ven afectadas por una actualización.

## Comunicación y encriptación

Dbvisit Replicate utiliza un canal de transporte seguro para enviar datos desde el servidor de origen al servidor de reporting (destino). Todos los comandos de red entre los componentes de Dbvisit Replicate (por ejemplo *fetcher*, *mine*, *apply* y *console*) se envían cifrados, con la clave que proporciona la autenticación y el cifrado.

Dbvisit Replicate ha sido diseñado y construido para trabajar, de forma nativa, en un entorno basado en *cloud*. Dbvisit Replicate permite a los usuarios de bases de datos Oracle crear y administrar fácilmente bases de datos de *reporting on premise* o en *Cloud*.

Dbvisit también ofrece un entorno de laboratorio basado en *cloud* que ofrece a los administradores acceso para explorar las funciones, la facilidad de uso y el poder de los productos.

## Conclusión

La creciente demanda de información en tiempo real requiere que las compañías desarrollen soluciones que proporcionen acceso a los datos de producción, sin afectar la disponibilidad, el rendimiento o la capacidad de respuesta de los sistemas en línea.

Las bases de datos de específicas para *reporting* están diseñadas para proporcionar los datos necesarios para *el reporting* en un entorno independiente de los repositorios de datos transaccionales en línea. Esto garantiza que el *reporting* no afecte al rendimiento de la base de datos o a los sistemas de proceso de transacciones en línea.

Existe una complejidad significativa en el mantenimiento de información actualizada en la base de datos de *reporting* sin afectar a la disponibilidad de los sistemas transaccionales. Aunque hay una gran variedad de opciones disponibles, solo la replicación de datos lógicos cumple con la amplia variedad de necesidades de una solución de este tipo:

- Carga muy baja en la base de datos fuente y en el entorno del servidor
- Replicación, casi en tiempo real, de los cambios en los datos, en la base de *reporting* de destino
- Acceso ininterrumpido al sistema fuente para usuarios en línea y otra actividades transaccionales

Dbvisit Replicate es una solución que cumple con estas necesidades, lo que permite a las compañías instalar bases de datos de *reporting* fiables, seguras, robustas y rentables en un entorno operativo heterogéneo. Dbvisit Replicate permite al administrador controlar los datos que se replican en la base de datos de *reporting*, incluyendo tablas, columnas e incluso filas basándose en criterios de coincidencia de datos. Dbvisit Replicate proporciona numerosas funciones para detectar y resolver conflictos en los datos.

Además de operar en una red corporativa, Dbvisit Replicate puede operar en un entorno en el que la base de datos de *reporting* esté alojada en un servicio en *cloud*, como Amazon Web Services (AWS). Este enfoque permite a las compañías obtener más beneficios de las bases de datos de *reporting*, incluyendo rentabilidad, escalabilidad, rendimiento y accesibilidad.

Para más información, contactar con Dbvisit Software en [sales@dbvisit.com](mailto:sales@dbvisit.com)



DATABASE TECHNOLOGY

Dbvisit Software Limited

[www.dbvisit.com](http://www.dbvisit.com)

[info@dbvisit.com](mailto:info@dbvisit.com)

EE.UU.: 1-800-933-8007

Resto del mundo: +64 9 950 3301

Fax: +64 9 950 3302

Dbvisit Software Limited

Apartado de correos 48180, Blockhouse Bay

Auckland 0604, Nueva Zelanda



**ds data systems**

[www.dsdata.es](http://www.dsdata.es)

[spssoftware@dsdata.es](mailto:spssoftware@dsdata.es)

Teléfono: + 34 91 579 66 46

C/ Arenal de Maudes, 10, 3º

28036 Madrid ( España)