

Documento técnico de Oracle
Septiembre de 2009

Compresión avanzada con Oracle Database 11g versión 2

Introducción.....	3
Oracle Advanced Compression	4
Compresión de datos de tablas	4
Compresión de tabla OLTP	4
Compresión de datos de archivos	7
Desduplicación de SecureFiles.....	7
Compresión de SecureFiles	8
Compresión de datos de backup	9
Compresión de Recovery Manager	9
Compresión de Data Pump	9
Compresión del tráfico de red.....	10
Conclusión	11

Introducción

Las empresas están experimentando un estallido en cuanto al volumen de datos necesario para llevar adelante sus negocios. Esta tendencia en el aumento de datos puede atribuirse a varios factores clave. Cambios recientes en cuanto a la regulación, como las leyes Sarbanes-Oxley e HIPAA, contribuyen a esta tendencia exigiéndoles a las empresas que retengan una gran cantidad de información por largos períodos de tiempo. La distribución masiva a través de Internet de contenidos ricos y de multimedia que ahora es posible gracias a los avances en la tecnología de banda ancha también contribuye al crecimiento general en el volumen de datos. La tendencia exponencial en el crecimiento de datos se ve exacerbada aún más con el advenimiento de la Web 2.0, con las aplicaciones colaborativas que promueven cantidades enormes de contenidos generados por los usuarios. Varios cálculos indican que el volumen de datos se duplica cada 2 a 3 años.

Este estallido repentino en el volumen de datos les presenta un desafío administrativo de enormes proporciones a los administradores de IT. El primero y principal es el costo en alza constante del almacenamiento: a pesar de que el costo de almacenamiento por MB ha declinado significativamente en los últimos años, el crecimiento en el volumen de datos que es necesario mantener constituye uno de los mayores costos en la mayoría de los presupuestos de IT. Además, el rendimiento y la escalabilidad de las aplicaciones deben seguir satisfaciendo las demandas de la empresa, incluso a medida que el volumen de datos estalla.

Oracle Database 11g introduce la opción Advanced Compression para ayudar a los clientes a sobrellevar estos desafíos. Las innovaciones en las tecnologías de compresión de Oracle ayudan a los clientes a reducir los recursos y costos que conlleva el administrar grandes volúmenes de datos. La introducción de estas atractivas nuevas tecnologías llega en el momento oportuno, ya que las bases de datos de tamaño de terabytes que se consideraban como una novedad ahora son moneda corriente en los centros de datos de las empresas.

Oracle Advanced Compression

La opción Advanced Compression de Oracle Database 11g introduce un conjunto integral de capacidades de compresión para ayudar a los clientes a maximizar la utilización de sus recursos y reducir costos. Les permite a los administradores de IT reducir el espacio de utilización general del almacenamiento de su base de datos al facilitar la compresión de todo tipo de datos, sean relacionales (tablas), sin estructurar (archivos) o datos de backup. A pesar de que por lo general se considera al ahorro en cuanto al almacenamiento como el beneficio más tangible de la compresión, las innovadoras tecnologías que se incluyen en la opción Advanced Compression están diseñadas para reducir los requisitos de recursos y los costos de tecnología para todos los componentes de su infraestructura de IT, incluidas la memoria y el ancho de banda de la red.

Compresión de datos de tablas

Oracle ha sido pionera en la introducción de la tecnología de compresión de bases de datos. Oracle Database 9i incorporó la compresión básica de tabla varios años atrás, opción que permitía comprimir datos cargados en operaciones de carga en bloque. Oracle Database 11g versión 1 presentó una nueva característica denominada Compresión de Tabla OLTP que permite la compresión de datos durante todo tipo de operaciones de manipulación de datos, incluidas las DML como INSERTAR y ACTUALIZAR. Además, la compresión de tabla OLTP reduce la sobrecarga de compresión asociada a las operaciones de escritura, por lo que también resulta apropiada para entornos transaccionales u OLTP. La compresión de tabla OLTP, por lo tanto, extiende los beneficios de la compresión a todas las cargas de trabajo de las aplicaciones.

Cabe destacar que la compresión básica de tabla es una característica base de Oracle Database 11g Enterprise Edition (EE). La compresión de tabla OLTP forma parte de la opción Oracle Advanced Compression, cuya licencia debe adquirirse por separado de Enterprise Edition.

Compresión de tabla OLTP

La opción de compresión de tabla OLTP de Oracle se vale de un algoritmo de compresión único que se diseñó específicamente para trabajar con aplicaciones OLTP. El algoritmo elimina valores duplicados dentro de un bloque de una base de datos, incluso a través de múltiples columnas. Los bloques comprimidos contienen una estructura denominada tabla de símbolo que mantiene metadatos de compresión. Al comprimirse un bloque, se eliminan los valores duplicados agregando primero una sola copia del valor duplicado a la tabla de símbolo. Luego se reemplaza cada valor duplicado por una corta referencia a la entrada adecuada en la tabla de símbolo. Mediante este innovador diseño, los datos comprimidos se encuentran autocontenidos en el bloque de la base de datos, ya que los metadatos utilizados para traducir los datos comprimidos a su estado original se encuentran almacenados dentro del mismo bloque. En comparación con los algoritmos de compresión de los competidores que mantienen una tabla de símbolos global, el enfoque único de Oracle ofrece significativos beneficios de rendimiento al no introducir I/O adicional al acceder a datos comprimidos.

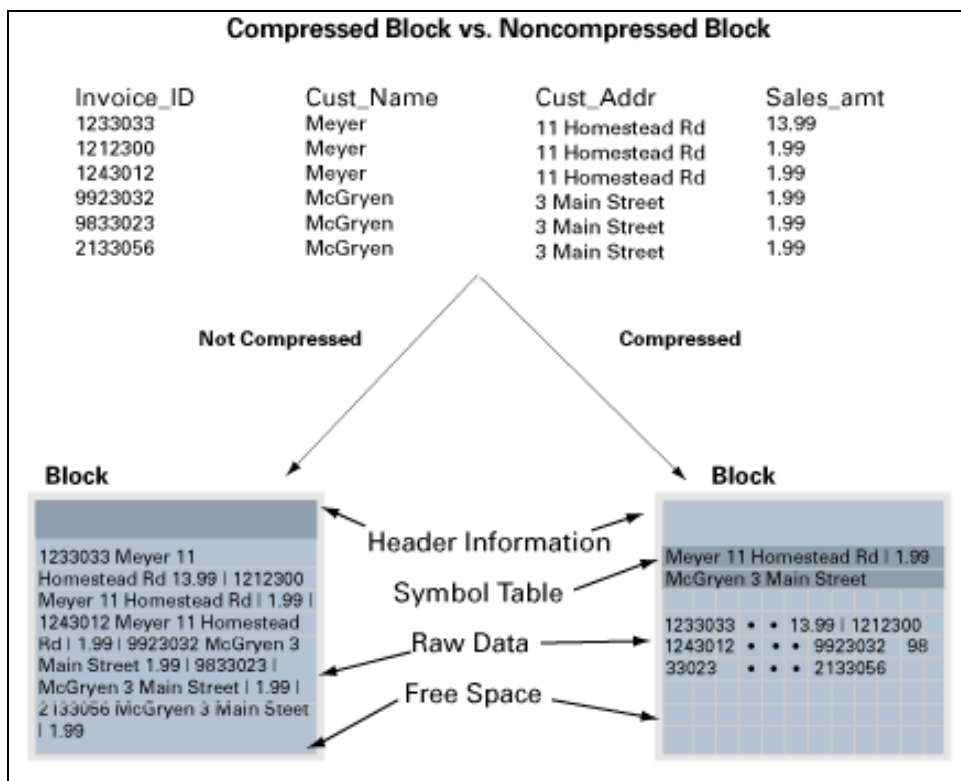


Figura 1: Bloque comprimido frente a bloque sin comprimir

Beneficios de la compresión de tabla OLTP

El índice de compresión que se logra en un entorno determinado depende del tipo de datos que se comprimirá, específicamente la cardinalidad de dichos datos. En general, los clientes pueden esperar una reducción en el consumo de espacio de almacenamiento de 2 a 3 veces mediante la utilización de la característica de compresión de tabla. Es decir, la cantidad de espacio utilizado por los datos sin comprimir será de dos a tres veces mayor que la de los datos comprimidos. Los beneficios de la compresión de tabla OLTP van más allá del ahorro en el espacio de almacenamiento en el disco. Una ventaja significativa es la habilidad de Oracle de leer bloques comprimidos directamente sin tener que primero descomprimir el bloque. Por lo tanto, no existe degradación de rendimiento medible al acceder a los datos comprimidos. De hecho, en muchos casos el rendimiento hasta podría mejorar debido a la reducción de I/O, ya que Oracle tendrá que acceder a menos bloques. Además, la caché de búfer resultará más eficaz mediante el almacenamiento de más datos sin tener que incrementar el espacio de utilización en la memoria.

Sobrecarga de rendimiento mínima

Como se mencionó antes, la compresión de tabla OLTP no produce ningún impacto adverso en las operaciones de lectura. Se lleva a cabo trabajo adicional mientras se escriben los datos, lo que

hace imposible que se elimine la sobrecarga en el rendimiento en las operaciones de escritura. Sin embargo, Oracle se ha esforzado para reducir al mínimo esa sobrecarga en la opción de compresión de tabla OLTP. Oracle comprime bloques en modo de proceso por lotes en vez de comprimir datos cada vez que una operación de lectura se lleva a cabo. Un bloque recién inicializado no se comprime hasta que los datos en el bloque alcancen un límite controlado internamente. Cuando una transacción hace que los datos en el bloque alcancen dicho límite, se comprimen todos los contenidos del bloque. Posteriormente, a medida que se añaden más datos al bloque y se alcanza el límite una vez más, se vuelve a comprimir todo el bloque para lograr el más alto nivel de compresión. Este proceso se repite hasta que Oracle determina que el bloque ya no se puede comprimir más. Sólo las transacciones que provocan la compresión del bloque experimentarán una pequeña sobrecarga de compresión. Por lo tanto, la mayoría de las transacciones OLTP en bloques comprimidos tendrá el mismo rendimiento que tendría en bloques sin comprimir.

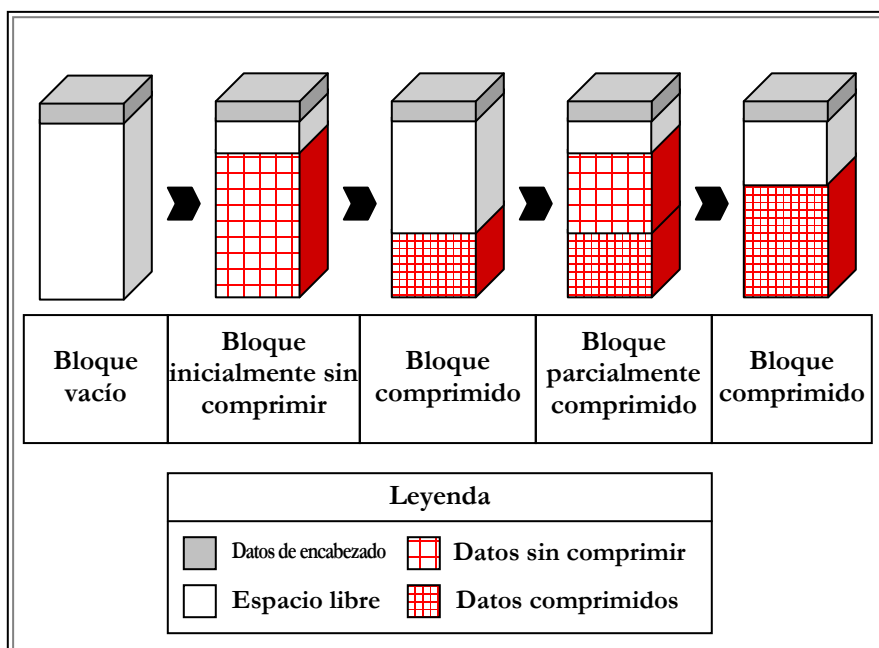


Figura 2: Proceso de compresión de tabla OLTP

Sintaxis de la compresión de tabla OLTP

```
CREATE TABLE emp (
    emp_id NUMBER
, first_name VARCHAR2(128)
, last_name VARCHAR2(128)
) COMPRESS FOR OLTP;
```

Compresión de datos de archivos

SecureFiles, una nueva característica de Oracle Database 11g, ofrece una arquitectura con todas las ventajas para el almacenamiento de contenidos sin estructurar, como pueden ser documentos, planillas de cálculo y archivos XML. SecureFiles se diseñó específicamente para brindar alto rendimiento para los datos de archivos, comparable al de los sistemas de archivo tradicionales pero reteniendo las ventajas de Oracle Database. SecureFiles se diseñó como reemplazo del LOB ANSI estándar y ofrece una fácil migración de los viejos archivos LOB de BasicFiles, el precursor de SecureFiles. Con SecureFiles, las organizaciones ahora pueden administrar todos los datos relacionales y datos de archivos asociados en Oracle utilizando un único modelo de seguridad/auditoría, un proceso unificado de backup y recuperación, y llevar a cabo una recuperación a través de toda la información. La opción Advanced Compression de Oracle Database 11g incluye tecnologías que reducen drásticamente el espacio necesario para el almacenamiento de datos de SecureFiles.

Desduplicación de SecureFiles

Es muy común que las aplicaciones almacenen réplicas exactas de archivos. Un ejemplo típico es una aplicación de correo electrónico en la que diversos usuarios reciben el mismo documento adjunto. La desduplicación de SecureFiles que incluye Advanced Compression es una tecnología inteligente que elimina copias duplicadas de datos de SecureFiles. Oracle almacena una imagen de los datos de SecureFiles y reemplaza las copias duplicadas con referencias a esta imagen. Tomemos como ejemplo una aplicación de correo electrónico donde 10 usuarios reciben un correo electrónico con el mismo adjunto de 1 MB. Sin la desduplicación de SecureFiles, el sistema almacenaría una copia del archivo por cada uno de los 10 usuarios, cuyo almacenamiento utilizaría 10 MB. Siguiendo nuestro ejemplo, si la aplicación de correo electrónico hubiera utilizado SecureFiles con desduplicación, hubiera almacenado el archivo de 1 MB sólo una vez. Esto resulta en un ahorro del 90% del almacenamiento necesario. Además del ahorro en cuanto a almacenamiento, la desduplicación de SecureFiles también aumenta el rendimiento de la aplicación. Específicamente, las operaciones de lectura y copia se vuelven más eficientes ya que sólo se escriben las referencias a la imagen de SecureFiles. Además, puede que las operaciones de lectura hasta mejoren si es que existen datos de SecureFiles duplicados en la caché de búfer.

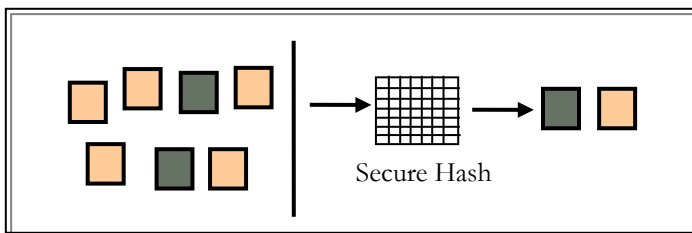


Figura 3: Desduplicación de SecureFiles

Sintaxis de la deduplicación de SecureFiles

```
CREATE TABLE images (
  image_id NUMBER,
  image BLOB)
  LOB(image) STORE AS SECUREFILE
  (TABLESPACE lob_tbs DEDUPLICATE);
```

Compresión de SecureFiles

La opción Advanced Compression de Oracle Database 11g proporciona aún otro mecanismo más para controlar el tamaño de sus datos de SecureFiles. Además de la deduplicación analizada anteriormente, la compresión de SecureFiles utiliza algoritmos de compresión estándar para minimizar aún más el espacio que se necesita para almacenar los datos de SecureFiles. Con la compresión de SecureFiles, los archivos típicos, como documentos o archivos XML, experimentan una reducción de 2 a 3 veces en tamaño. Utilizando inteligencia incorporada, la compresión de SecureFiles evita automáticamente la compresión de datos que no se podrían comprimir más, por ejemplo, el caso de un documento que ya fue comprimido por la herramienta de un tercero antes de ser insertado en la base de datos como un archivo de SecureFiles. Las aplicaciones aún pueden llevar a cabo operaciones de lectura y escritura al azar en datos de SecureFiles comprimidos ya que los datos comprimidos se dividen en pedazos de datos más pequeños. Esto puede mejorar el rendimiento infinitamente cuando se lo compara con la compresión de archivos enteros antes de insertarlos en la base de datos.

Existen tres niveles de compresión de SecureFiles: BAJO, MEDIO y ALTO. En forma predeterminada, la compresión de SecureFiles utiliza el nivel MEDIO, el cual suele brindar una buena compresión con una escasa sobrecarga del CPU: entre 3 y 5%. El nivel BAJO de compresión de SecureFiles, novedad en Oracle Database 11g versión 2, está optimizado para lograr un alto rendimiento. El nivel BAJO de compresión de SecureFiles realmente conserva un 80% de la compresión lograda con el nivel MEDIO, aunque utiliza 3 veces menos recursos del CPU. Por último, el nivel ALTO de compresión de SecureFiles logra el más alto nivel de ahorro de almacenamiento, pero implica la mayor sobrecarga del CPU.

Sintaxis de la compresión de SecureFiles

```
CREATE TABLE images (
  image_id NUMBER,
  image BLOB)
  LOB(image) STORE AS SECUREFILE
  (TABLESPACE lob_tbs COMPRESS);
```


Compresión de datos de backup

Además de comprimir datos almacenados dentro de la base de datos, Oracle Advanced Compression también incluye la capacidad de comprimir datos de backup. Recovery Manager (RMAN) y Data Pump son dos de las herramientas más comúnmente utilizadas para realizar backup de los datos almacenados en Oracle Database. RMAN hace un backup bloque por bloque de los datos de la base de datos, también denominado backup "físico", que puede utilizarse para llevar a cabo la recuperación de la base de datos, del espacio de tabla o de bloques. Data Pump se utiliza para llevar a cabo un backup "lógico" descargando datos de una o más tablas a un archivo plano. Oracle Advanced Compression incluye la capacidad para comprimir los datos de backup generados por estas dos herramientas.

Compresión de Recovery Manager

El crecimiento continuo en cuanto a bases de datos empresariales supone un enorme desafío para los administradores de bases de datos. El tamaño de la base de datos tiene un impacto directo en el espacio necesario para almacenar los backups de la base de datos y el rendimiento de los procesos de backup. Oracle Advanced Compression incluye tecnología de compresión RMAN, que puede reducir drásticamente los requisitos de almacenamiento para datos de backup. Debido a la estrecha integración de RMAN con Oracle Database, los datos de backup se comprimen antes de ser escritos en el disco o cinta y no es necesario descomprimirlos antes de ser recuperados, lo que proporciona una enorme reducción en los costos de almacenamiento.

Existen tres niveles de compresión de RMAN: BAJO, MEDIO y ALTO. La cantidad de ahorro de espacio de almacenamiento aumenta del nivel BAJO al ALTO, a la vez que posiblemente se consuman más recursos del CPU.

Sintaxis para configurar el algoritmo de compresión de RMAN:

```
RMAN> SET COMPRESSION ALGORITHM 'LOW|MEDIUM|HIGH';
```

Sintaxis para realizar un backup comprimido con RMAN:

```
RMAN> backup as COMPRESSED BACKUPSET database archivelog all;
```

Compresión de Data Pump

La habilidad de comprimir los metadatos asociados con una operación de Data Pump se introdujo por primera vez en Oracle Database 10g versión 2. En Oracle Database 11g, esta capacidad de compresión ha sido extendida para que se puedan comprimir los datos de una tabla al exportarla. La compresión de Data Pump es una operación en línea, por lo tanto la reducción

en el tamaño del archivo de volcado de memoria resulta en un gran ahorro en el espacio de disco utilizado. A diferencia de las herramientas de compresión de los sistemas operativos o sistemas de archivos, la compresión de Data Pump se realiza totalmente en línea también en cuanto a la importación, por lo tanto no es necesario descomprimir un archivo de volcado de memoria antes de importarlo. Se descomprime el conjunto de archivos de volcado de memoria automáticamente durante una importación sin que el administrador de la base de datos haga nada más.

En el siguiente ejemplo de compresión de la base de datos de muestra de Oracle, los esquemas OE y SH se importaron mientras que simultáneamente se comprimieron todos los datos y metadatos. El tamaño del archivo de volcado de memoria se redujo en un 74,67%.

Se utilizaron tres versiones de la herramienta gzip (GNU zip) y una herramienta de compresión UNIX para comprimir el conjunto de archivos de volcado de memoria de 6,0 MB. La reducción en el tamaño del archivo de volcado de memoria fue comparable a la de la compresión de Data Pump. Cabe notar que la reducción en el tamaño del archivo de volcado de memoria variará según el tipo de dato y otros factores.

La funcionalidad de Data Pump se encuentra disponible en su totalidad al utilizarse un archivo comprimido. Cualquier comando que se utilice en un archivo normal también funcionará en un archivo comprimido. Los usuarios cuentan con las opciones que se describen a continuación para determinar qué partes de un conjunto de archivos de volcado de memoria deben comprimirse:

- **ALL** permite la compresión de la operación de exportación en su totalidad.
- **DATA-ONLY** escribe todos los datos en el archivo de volcado de memoria en formato comprimido.
- **METADATA-ONLY** escribe todos los metadatos en el archivo de volcado de memoria en formato comprimido. Esto es lo que se encuentra predeterminado.
- **NONE** deshabilita la compresión de la operación de exportación en su totalidad.

Para obtener más información sobre Oracle Data Pump, por favor visite <http://www.oracle.com/technology/products/database/utilities/index.html>

Cómo habilitar la compresión de Data Pump

El ejemplo que se encuentra a continuación muestra cómo se habilita la compresión de Data Pump:

```
expdp hr FULL=y DUMPFILE=dpump_dir:full.dmp COMPRESS;
```

Compresión del tráfico de red

Data Guard proporciona la infraestructura de software de administración, control y automatización para crear, mantener y controlar una o más bases de datos standby y así proteger

los datos de las empresas de fallas, desastres, errores y corrupción. Data Guard mantiene la sincronización de las bases de datos primarias y standby utilizando datos redo (la información necesaria para recuperar una transacción). A medida que se llevan a cabo transacciones en la base de datos primaria, se generan y escriben datos redo en los archivos de registro redo locales. Se utiliza Data Guard Redo Transport Services para transferir estos datos redo a los sitios standby. Con Advanced Compression, es posible transmitir los datos redo en un formato comprimido a fin de reducir el consumo del ancho de banda y en algunos casos acortar el tiempo de transmisión de esos datos. A partir de Oracle Database 11g versión 2, los datos redo pueden transmitirse en un formato comprimido cuando la configuración de Oracle Data Guard usa transporte sincrónico de datos redo (SYNC), o bien transporte asincrónico de datos redo (ASYNC).

Para obtener más información sobre Oracle Data Guard, por favor visite <http://www.oracle.com/technology/deploy/availability/index.html>

Conclusión

El estallido en el volumen de datos que experimentan las empresas trae aparejados importantes desafíos. Las empresas deben adaptarse rápidamente al entorno cambiante de sus negocios sin que esto tenga impacto en el resultado final. Los administradores de IT deben administrar eficientemente las infraestructuras existentes para controlar costos, pero a la vez seguir brindando un extraordinario rendimiento en las aplicaciones.

La opción Advanced Compression de Oracle Database 11g proporciona un conjunto robusto de capacidades que les permiten a los administradores de IT tener éxito en este complejo entorno. Aprovechando la opción Advanced Compression, las empresas pueden administrar eficientemente sus crecientes requisitos de datos a través de todos los componentes de sus centros de datos, minimizando costos y a la vez logrando el más alto rendimiento en sus aplicaciones.

ORACLE®

Compresión avanzada con
Oracle Database 11g versión 2
Septiembre de 2009
Autor: William Hodak
Colaboradores: Kevin Jernigan

Oracle Corporation
Sede mundial
500 Oracle Parkway
Redwood Shores, CA 94065
EE. UU.

Consultas internacionales:
Teléfono: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com



Oracle is committed to developing practices and products that help protect the environment

Copyright © 2009, Oracle o sus subsidiarias. Todos los derechos reservados. Este documento se proporciona exclusivamente con fines informativos y sus contenidos están sujetos a cambio sin previo aviso. No se garantiza que el presente documento esté libre de errores, ni tampoco está sujeto a otras garantías y condiciones, ya sean expresadas oralmente o implícitas en la legislación, incluidas garantías y condiciones implícitas de comerciabilidad o adecuación para un propósito en particular. Oracle se exime específicamente de cualquier responsabilidad respecto del presente documento y tampoco se establece relación contractual directa o indirecta alguna mediante el mismo. El presente documento no podrá reproducirse, ni transmitirse de forma ni por medio alguno, sea éste electrónico o mecánico, para finalidad alguna, sin el expreso consentimiento escrito de Oracle Corporation.

Oracle es una marca comercial registrada de Oracle Corporation y/o sus subsidiarias. Todos los demás nombres pueden ser marcas comerciales de sus respectivos propietarios.