

IBM SPSS Analytic Server
Versión 2.1

Guía del administrador

IBM

Nota

Antes de utilizar esta información y el producto que soporta, lea la información de “Avisos” en la página 13.

Información sobre el producto

Esta edición se aplica a la versión 2.1, release 1, modificación 0 de IBM SPSS Analytic Server y a todos los releases y modificaciones siguientes hasta que se indique lo contrario en nuevas ediciones.

Contenido

Capítulo 1. Descripción general	1	Recopilador de registros	11
Arquitectura	2	Problemas comunes	12
Novedades para los administradores en la versión 2.1	2		
Capítulo 2. Gestión de inquilinos	3	Avisos	13
Reglas de denominación (consola de	4	Marcas registradas	15
Capítulo 3. Iniciación de usuarios.	7		
Capítulo 4. Nombres de trabajos de Analytic Server.	9		
Capítulo 5. Resolución de problemas	11		
Registro	11		
Información sobre la versión.	11		

Capítulo 1. Descripción general

IBM® SPSS Analytic Server es una solución de análisis masivo de datos que combina tecnología de IBM SPSS con sistemas de datos masivos, y que permite trabajar con interfaces de usuario de IBM SPSS conocidas para resolver problemas a una escala antes impensable.

Por qué es importante el análisis masivo de datos

El volumen de datos recopilados por las organizaciones crece de forma exponencial; por ejemplo, las empresas financieras y de venta al por menor tienen todas las transacciones de clientes de un año (o dos años, o diez), los proveedores de telecomunicaciones tienen los registros de datos de llamadas (CDR) y lecturas de sensores de dispositivos, y las empresas de internet tienen los resultados de los rastreos web.

Un análisis masivo de datos es necesario cuando existe:

- Un gran volumen de datos (terabytes, petabytes o exabytes), sobre todo cuando es una mezcla de datos estructurados y no estructurados.
- Datos que cambian/se acumulan con rapidez.

El análisis masivo de datos también es de ayuda cuando:

- Se construye un gran número de modelos (del orden de miles).
- Los modelos se construyen/renuevan con frecuencia.

Retos

Las mismas organizaciones que recopilan grandes volúmenes de datos suelen tener dificultades a la hora de utilizarlos, por una serie de razones:

- la arquitectura de los productos analíticos tradicionales no está pensada para la computación distribuida, y
- Los algoritmos estadísticos existentes no están diseñados para trabajar con cantidades masivas de datos (tales algoritmos esperan que los datos les lleguen, pero cuesta mucho mover datos masivos), por tanto
- el análisis de datos masivos con tecnología puntera requiere nuevas habilidades y un conocimiento a fondo de los sistemas de datos masivos. Muy pocos analistas poseen estas habilidades.
- Las soluciones residentes en memoria son aptas para problemas de tamaño medio, pero no escalan bien a datos realmente masivos.

Solución

Analytic Server proporciona:

- Una arquitectura centrada en datos que saca partido de sistemas de datos masivos tales como Hadoop Map/Reduce con datos en HDFS.
- Una interfaz definida para incorporar nuevos algoritmos estadísticos diseñados para ir a los datos.
- Conocidas interfaces de usuario de IBM SPSS que ocultan los detalles de los entornos de datos masivos, de modo que el analista pueda centrarse en el análisis de los datos.
- Una solución escalable a problemas de cualquier tamaño.

Arquitectura

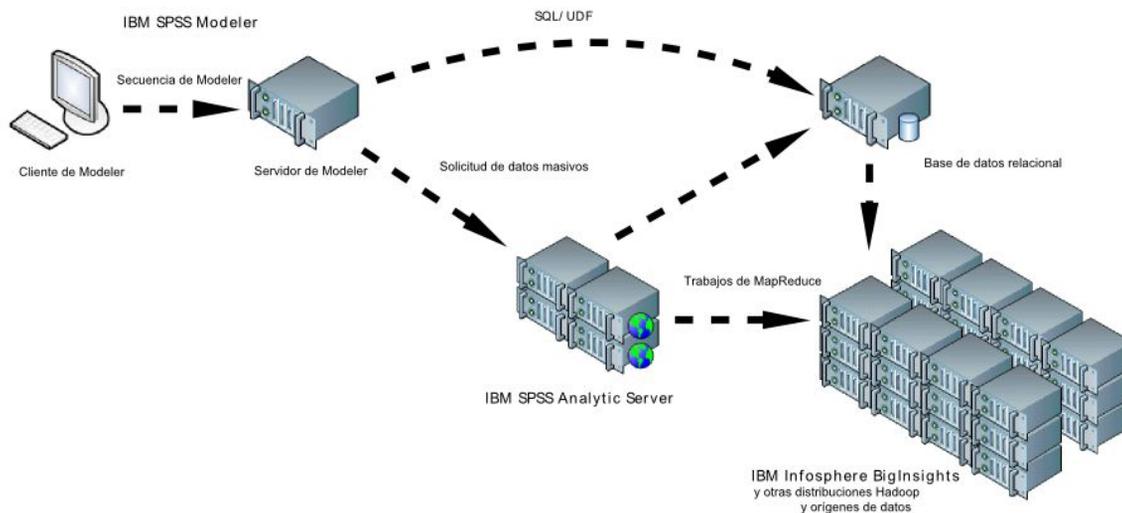


Figura 1. Arquitectura

Analytic Server se sitúa entre una aplicación cliente y una nube Hadoop. Suponiendo que los datos residan en la nube, la forma de trabajar con Analytic Server sería a grandes rasgos:

1. Se definen los orígenes de datos de Analytic Server que dan acceso a los datos de la nube.
2. Se define el análisis que se desea realizar en la aplicación cliente. En el release actual, la aplicación de cliente es IBM SPSS Modeler.
3. Cuando se ejecuta el análisis, la aplicación cliente envía una solicitud de ejecución de Analytic Server.
4. Analytic Server organiza el trabajo para que ejecute en la nube de Hadoop e informa de los resultados a la aplicación cliente.
5. Los resultados pueden utilizarse para definir análisis adicionales, con lo que se repetiría el ciclo.

Novedades para los administradores en la versión 2.1

Consola de Analytic Server

Rol de lector

Dentro de un inquilino, puede asignar un rol de lector a los usuarios y grupos que no pueden iniciar sesión en Analytic Server, pero pueden leer los orígenes de datos de Analytic Server a través del nodo de origen de Analytic Server en Modeler.

Capítulo 2. Gestión de inquilinos

Los inquilinos proporcionan una división de alto nivel de usuarios, proyectos y orígenes de datos, de modo que los objetos no se pueden compartir entre inquilinos. Cada usuario accede al sistema en el contexto del inquilino que le ha sido asignado.

En la consola de Analytic Server puede gestionar los inquilinos y asignar usuarios a los inquilinos. La vista de la página Inquilinos depende del rol del usuario que ha iniciado sesión en la consola:

- El administrador "superusuario" que se configura durante la instalación es el administrador de los inquilinos. Este usuario es el único que puede crear inquilinos y editar las propiedades de cualquiera de ellos.
- Los usuarios con el rol de administrador pueden editar las propiedades del inquilino con el que hayan iniciado sesión.
- Los usuarios con el rol de usuario no pueden editar las propiedades de ningún inquilino. La página Inquilinos se les oculta.
- Los usuarios con el rol de lector no pueden editar los orígenes de datos ni incluso iniciar sesión en la consola de Analytic Server.

Los administradores pueden acceder a las páginas Proyectos y Orígenes de datos, y gestionar los proyectos y orígenes de datos para administrarlos o realizar una limpieza. Consulte la publicación *IBM SPSS Analytic Server Guía del usuario* para obtener más información.

Lista de inquilinos

La página principal Inquilinos muestra los inquilinos existentes en una tabla. Solo el administrador "superusuario" puede realizar ediciones en esta página.

- Pulse el nombre de un inquilino para visualizar sus detalles y editar sus propiedades.
- Pulse el URL de un inquilino para abrir la consola en el contexto de dicho inquilino.

Nota: Se finalizará su sesión y deberá iniciar sesión con las credenciales válidas para dicho inquilino.

- Escriba en el área de búsqueda para filtrar la lista de forma que solo se muestren los inquilinos que tienen la cadena de búsqueda en su nombre.
- Pulse **Nuevo** para crear un inquilino con el nombre que especifique en el diálogo **Añadir inquilino nuevo**. Consulte la sección "Reglas de denominación (consola de)" en la página 4 para obtener información acerca de los nombres que puede asignar a los inquilinos.
- Pulse **Suprimir** para eliminar el inquilino.
- Pulse **Renovar** para actualizar la lista.

Detalles del inquilino individual

El área de contenido se divide en varias secciones contraíbles.

Detalles

Nombre

Campo de texto editable que muestra el nombre del inquilino.

Descripción

Campo de texto editable que permite facilitar un texto explicativo sobre el inquilino.

URL

Es el URL que se facilita a los usuarios para que inicien sesión en el inquilino a través de

la consola de Analytic Server, y para configurar el servidor de SPSS Modeler. Consulte *IBM SPSS Analytic Server Guía de instalación y configuración* para obtener detalles acerca de cómo configurar SPSS Modeler.

Estado

Hay inquilinos **activos** actualmente en uso. Cambiar un usuario a **Inactivo** impide que los usuarios inicien sesión en ese inquilino, pero no suprime ningún dato subyacente.

Principales

Los principales son usuarios y grupos obtenidos del proveedor de seguridad configurado durante la instalación. Pueden añadirse principales a un inquilino en calidad de administradores o usuarios.

- Al escribirse en el cuadro de texto, se filtran los usuarios y grupos que tengan la cadena de búsqueda en el nombre. Seleccione **Administrador**, **Usuario** o **Lector** en la lista desplegable para asignarles el rol en el inquilino. Pulse **Añadir participante** para añadirlo a la lista de autores.
- Para eliminar un participante, seleccione un usuario o grupo en la lista de miembros y pulse **Eliminar participante**.

Métricas

Le permite configurar los límites de recursos para un inquilino. Informa acerca del espacio de disco que utiliza actualmente el inquilino.

- Puede establecer una cuota máxima de espacio de disco para el inquilino. Cuando se alcanza este límite, no se puede grabar nada más en disco para este inquilino hasta que se haya borrado el espacio de disco suficiente y el uso de espacio de disco por parte del inquilino se encuentre por debajo de la cuota.
- Puede establecer una cuota máxima de espacio de disco para el inquilino. Cuando se supera este límite, los principales de este inquilino no pueden enviar ningún trabajo analítico hasta que se haya borrado el espacio de disco suficiente y el uso de espacio de disco por parte del inquilino se encuentre por debajo de la cuota.
- Puede establecer una cuota máxima de trabajos paralelos que pueden ejecutarse en este inquilino de una sola vez. Cuando se supera esta cuota, los principales no pueden enviar ningún trabajo analítico en este inquilino hasta que se haya completado el trabajo que está en ejecución.
- Puede establecer el número máximo de campos que puede tener un origen de datos. El límite se comprueba siempre que se crea o actualiza un origen de datos.
- Puede establecer el tamaño máximo en MB. El límite se comprueba al cargar un archivo.

Configuración del proveedor de seguridad

Le permite especificar el proveedor de autenticación de usuarios. **Valor predeterminado** utiliza el proveedor del inquilino predeterminado, el cual se establece durante la instalación y configuración. **LDAP** permite autenticar usuarios con un servidor LDAP externo, tal como Active Directory u OpenLDAP. Especifique uno de los valores para el proveedor y, opcionalmente, especifique valores de filtro para controlar los usuarios y grupos disponibles en la sección Principales.

Reglas de denominación (consola de

Para todo aquello a lo que se pueda asignar un nombre exclusivo en Analytic Server, por ejemplo, los orígenes de datos y proyectos, se aplican las siguientes reglas para los nombres.

- En un solo inquilino, los nombres deben ser exclusivos en los objetos del mismo tipo. Por ejemplo, dos orígenes de datos no pueden tener el nombre insuranceClaims, pero un origen de datos y un proyecto pueden tener cada uno el nombre insuranceClaims.
- Los nombres son sensibles a las mayúsculas y minúsculas. Por ejemplo, insuranceClaims e InsuranceClaims se consideran nombres exclusivos.

- Los nombres omiten los espacios en blanco iniciales y finales.
- Los siguientes caracteres no son válidos en los nombres.
~, #, %, &, *, {, }, \\, :, <, >, ?, /, |, ", \t, \r, \n

Capítulo 3. Iniciación de usuarios

Indique a los usuarios que vayan a `http://<host>:<puerto>/<raíz-contexto>/admin/<inquilino>` y escriban su nombre de usuario y contraseña para iniciar la sesión en la consola de Analytic Server.

<host>

La dirección del host Analytic Server.

<puerto>

El puerto donde escucha Analytic Server. De forma predeterminada es 8080.

<raíz-contexto>

La raíz de contexto de Analytic Server. De forma predeterminada es `analyticserver`.

<inquilino>

En un entorno de varios inquilinos, el inquilino al que pertenece. En un entorno de un solo inquilino, el inquilino predeterminado es **ibm**.

Por ejemplo, si la máquina host tiene la dirección IP 9.86.44.232, ha creado un inquilino "mycompany" y ha añadido usuarios a este último y los demás valores se han dejado en sus valores predeterminados, los usuarios deben navegar a `http://9.86.44.232:8080/analyticserver/admin/mycompany` para acceder a la consola de Analytic Server.

Capítulo 4. Nombres de trabajos de Analytic Server

Analytic Server genera trabajos de reducción de correlaciones que se pueden supervisar mediante el rastreador de trabajos del clúster de Hadoop.

El nombre del trabajo de reducción de correlaciones tiene la estructura siguiente.

AS/{nombre inquilino}/{nombre usuario}/{nombre algoritmo}

{nombre inquilino}

Es el nombre del inquilino bajo el que se ejecuta el trabajo.

{nombre usuario}

Es el usuario que ha solicitado el trabajo.

{nombre algoritmo}

Es el algoritmo principal del trabajo. Tenga en cuenta que una sola corriente puede generar varios trabajos de reducción de correlaciones. Del mismo modo, si hay varias operaciones en una corriente, éstas pueden estar contenidas en un solo trabajo de reducción de correlaciones.

Capítulo 5. Resolución de problemas

Analytic Server proporciona varias herramientas útiles para la determinación de problemas.

Registro

Analytic Server crea archivos de rastreo y archivos de registro de cliente que son de utilidad a la hora de diagnosticar un problema. Con la instalación predeterminada de Liberty, puede encontrar los archivos de registro en el directorio `{AS_ROOT}/ae_wlpserver/usr/servers/aeserver/logs`.

La configuración de registro predeterminada produce dos archivos de registro que se renuevan a diario.

as.log Este archivo contiene el resumen de alto nivel de los mensajes informativos de error y aviso. Cuando se produzca un error de servidor que no pueda resolverse a partir del mensaje de error mostrado en la interfaz de usuario, empiece por consultar este archivo.

as_trace.log Este archivo contiene todas las entradas de `ae.log`, e información adicional destinada principalmente al soporte y desarrollo de IBM a efectos de depuración.

Analytic Server utiliza por debajo Apache LOG4J como recurso de registro. Con `log4j`, el registro puede ajustarse de forma dinámica editando el archivo de configuración `{AS_SERVER_ROOT}/configuration/log4j.xml`. Es posible que el soporte de IBM le pida esto para ayudarle a diagnosticar un problema, o puede que le interese modificarlo para limitar el número de archivos de registro que se mantienen. Los cambios en el archivo se detectan automáticamente en unos segundos, de modo que no es necesario reiniciar Analytic Server.

Para obtener información adicional relativa a `log4j` y al archivo de configuración, consulte la documentación en el sitio oficial de Apache <http://logging.apache.org/log4j/>.

Información sobre la versión

Puede determinarse la versión instalada de Analytic Server consultando la carpeta `{RAÍZ_AS}/properties/version`. Los siguientes archivos contienen información de versión.

IBM_SPSS_Analytic_Server-*.swtag
Contiene información de producto detallada.

version.txt
Versión y número de compilación del producto instalado.

Recopilador de registros

Cuando no se pueden resolver los problemas consultando directamente los archivos de registro, puede empaquetar todos los registros y enviarlos al soporte de IBM. Se proporciona un programa de utilidad para simplificar la recopilación de todos los datos necesarios.

Utilizando un shell de mandatos, ejecute los siguientes mandatos:

```
cd {AS_ROOT}/bin
run >sh ./logcollector.sh
```

Estos mandatos crean un archivo comprimido bajo `{AS_ROOT}/bin`. El archivo comprimido contiene todos los archivos de registro y la información de versión del producto.

Problemas comunes

En esta sección se describen algunos problemas de administración comunes y cómo puede arreglarlos.

Seguridad

La autenticación Kerberos falla al intentar acceder al origen de datos de HCatalog

Si aparecen errores en el registro como el siguiente:

```
cause:javax.security.sasl.SaslException: error al inicializar el contexto de seguridad
```

```
com.spss.analyticframework.api.exceptions.ComponentException: no se puede acceder a HCatalog
```

Debe asegurarse de que el tíquet de otorgamiento de tíquet Kerberos de usuario de HDFS está almacenado en memoria caché y disponible en el host de servidor de Analytic Server. Para hacer esto:

1. Detenga el proceso de Analytic Server.
2. Ejecute `kinit -f $hdfs.user` en el host de Analytic Server, donde **\$hdfs.user** es como está definido en el archivo `config.properties` y tiene permiso de escritura para la raíz de análisis.
3. Inicie Analytic Server.

Consola de Analytic Server

Acceso a la consola de Analytic Server desde Safari on iOS

La lista desplegable de inquilinos no funciona y no puede actualizar el modelo de datos de los orígenes de datos basados en archivo. Utilice otro navegador al realizar estas acciones.

Ejecutando corrientes

La ejecución de los trabajos PySpark ha fallado

1. En la consola Ambari, añadir `export SPARK_HOME=/usr/iop/current/spark-client` al parámetro **yarn-env** del servicio YARN.
2. Reiniciar el servicio YARN.

Avisos

Esta información se ha desarrollado para productos y servicios que se ofrecen en los EE.UU.

Es posible que IBM no ofrezca en otros países los productos, servicios o características que se describen en este documento. Póngase en contacto con el representante local de IBM si desea obtener información sobre los productos y servicios disponibles en su zona. Cualquier referencia hecha en esta publicación a un producto, programa o servicio de IBM no pretende indicar ni implica que solo pueda utilizarse ese producto, programa o servicio de IBM. En su lugar, se puede utilizar cualquier producto, programa o servicio equivalente que no infrinja ninguno de los derechos de propiedad intelectual de IBM. No obstante, es responsabilidad del usuario evaluar y verificar el funcionamiento de cualquier producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patente pendientes que cubran la materia descrita en este documento. El suministro de este documento no le otorga ninguna licencia sobre estas patentes. Puede enviar consultas sobre licencias, por escrito, a:

IBM Director of Licensing
IBM Corporation
North Castle
Drive
Armonk, NY 10504-1785
EE.UU.

Para consultas sobre licencias relacionadas con información de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus consultas, por escrito, a:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japón

El párrafo siguiente no se aplica al Reino Unido ni a ningún otro país en que dichas disposiciones entren en contradicción con las leyes locales: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍA DE NINGÚN TIPO, NI EXPLÍCITA NI IMPLÍCITA, INCLUYENDO, PERO NO LIMITÁNDOSE, A LAS GARANTÍAS IMPLÍCITAS DE NO VULNERABILIDAD, COMERCIALIZACIÓN O ADECUACIÓN A UN PROPÓSITO DETERMINADO. Algunas legislaciones no contemplan la declaración de limitación de responsabilidad, ni implícitas ni explícitas, en determinadas transacciones, por lo que cabe la posibilidad de que esta declaración no sea aplicable en su caso.

Esta información puede contener imprecisiones técnicas o errores tipográficos. Se realizan cambios periódicos en la información aquí contenida; estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar mejoras y/o cambios en los productos y/o programas descritos en esta publicación en cualquier momento y sin previo aviso.

Todas las referencias hechas en esta información a sitios Web que no son de IBM se proporcionan únicamente para su información y no representan en modo alguno una recomendación de dichos sitios Web. Los materiales de dichos sitios web no forman parte de los materiales para este producto de IBM y el uso de dichos sitios web corre a cuenta y riesgo del Cliente.

IBM puede utilizar o distribuir la información que el usuario le suministre en el modo que considere apropiado sin incurrir en ninguna obligación con el usuario.

Los propietarios de licencia de este programa que deseen tener información sobre el mismo con el fin de: (i) intercambiar información entre programas creados de forma independiente y otros programas (incluido éste) y (ii) utilizar mutuamente la información que se ha intercambiado, deberán ponerse en contacto con:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
EE.UU.

Dicha información puede estar disponible sujeta a los términos y condiciones adecuados, incluido, en algunos casos, el pago de una tasa.

El programa bajo licencia que se describe en este documento y todo el material bajo licencia disponible los proporciona IBM bajo los términos de las Condiciones Generales de IBM, Acuerdo Internacional de Programas Bajo Licencia de IBM o cualquier acuerdo equivalente entre las partes.

Los datos de rendimiento contenidos aquí se han determinado en un entorno controlado. Por lo tanto, los resultados obtenidos en otros entornos operativos podrían variar significativamente. Algunas mediciones pueden haberse realizado en sistemas de nivel de desarrollo y no hay garantía de que dichas mediciones sean las mismas en sistemas de disponibilidad general. Además, es posible que algunas mediciones se hayan estimado mediante extrapolación. Los resultados reales pueden variar. Los usuarios de esta documentación deberían verificar los datos aplicables a su entorno específico.

La información relativa a productos que no son de IBM la han proporcionado los proveedores de dichos productos, sus anuncios publicados u otras fuentes disponibles públicamente. IBM no ha probado esos productos y no puede confirmar la precisión del rendimiento, la compatibilidad ni ninguna otra afirmación relacionada con productos no IBM. Las preguntas sobre las prestaciones de los productos no IBM se deberán dirigir a los proveedores de estos productos.

Todas las declaraciones relativas a la orientación o intención futura de IBM están sujetas a cambio o anulación sin previo aviso y representan solamente metas y objetivos.

Todos los precios de IBM que se muestran son precios de venta al público sugeridos por IBM, son actuales y están sujetos a cambios sin previo aviso. Los precios de los proveedores pueden variar.

Esta información está pensada sólo a efectos de planificación. La información contenida en el presente documento está sujeta a cambios antes de que los productos descritos salgan al mercado.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos de la manera más completa posible, los ejemplos pueden incluir nombres de personas, compañías, marcas y productos. Todos estos nombres son ficticios y cualquier similitud a los nombres y direcciones que haya utilizado una empresa real es pura coincidencia.

Cada una de las copias, totales o parciales, de estos programas de ejemplo o cualquier trabajo derivado de ellos, debe incluir el siguiente aviso de copyright:

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos de la manera más completa posible, los ejemplos pueden incluir nombres de personas, compañías, marcas y productos. Todos estos nombres son ficticios y cualquier similitud a los nombres y direcciones que haya utilizado una empresa real es pura coincidencia.

Cada una de las copias, totales o parciales, de estos programas de ejemplo o cualquier trabajo derivado de ellos, debe incluir el siguiente aviso de copyright:

© (nombre de la empresa) (año). Partes de este código provienen de programas de ejemplo de IBM Corp.

© Copyright IBM Corp. _especifique el año o años_. Todos los derechos reservados.

Si está viendo esta información en copia software, es posible que las fotografías e ilustraciones a color no aparezcan.

Marcas registradas

IBM, el logotipo de IBM e ibm.com son marcas registradas o marcas comerciales registradas de International Business Machines Corp., registradas en muchas jurisdicciones en todo el mundo. Otros nombres de productos y servicios pueden ser marcas registradas de IBM y otras compañías. Está disponible una lista actual de marcas registradas de IBM en la web en "Copyright and trademark information" en www.ibm.com/legal/copytrade.shtml.

Adobe, el logotipo Adobe, PostScript y el logotipo PostScript son marcas registradas o marcas comerciales de Adobe Systems Incorporated en Estados Unidos y/u otros países.

IT Infrastructure Library es una marca registrada de Central Computer and Telecommunications Agency, que ahora forma parte de la Oficina Comercial del Gobierno (Office of Government Commerce).

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o de sus subsidiarias en los Estados Unidos y en otros países.

Linux es una marca registrada de Linus Torvalds en Estados Unidos o en otros países.

Microsoft, Windows, Windows NT y el logotipo de Windows son marcas registradas de Microsoft Corporation en los Estados Unidos o en otros países.

ITIL es una marca registrada y una marca de comunidad registrada de The Minister for the Cabinet Office y está registrada en la Oficina de marcas registradas y patentes de EE.UU. (U.S. Patent and Trademark Office).

UNIX es una marca registrada de The Open Group en Estados Unidos, otros países o en ambos.

Cell Broadband Engine es una marca comercial de Sony Computer Entertainment, Inc. en Estados Unidos, otros países o ambos y se utiliza bajo licencia.

Linear Tape-Open, LTO, el logotipo de LTO, Ultrium y el logotipo de Ultrium son marcas comerciales de HP, IBM Corp. y Quantum en Estados Unidos y otros países.



Impreso en España