

IBM SPSS Analytic Server
Versão 2.1

Guia do Administrador

IBM

Nota

Antes de utilizar estas informações e o produto suportado por elas, leia as informações em “Avisos” na página 13.

Informações sobre o Produto

Esta edição se aplica à versão do 2.1, liberação 1, modificação 0 do IBM SPSS Analytic Server e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

Índice

Capítulo 1. Visão Geral (Overview) . . .	1	Informações de versão	11
Arquitetura	2	Coletor de logs	11
O Que Há de Novo para Administradores na Versão		Problemas Comuns	12
2.1.	2		
Capítulo 2. Gerenciamento de		Avisos	13
Arrendatário	3	Marcas Registradas	15
Regras de nomenclatura	4		
Capítulo 3. Introdução para usuários ..	7		
Capítulo 4. Nomes de tarefa do Analytic			
Server	9		
Capítulo 5. Resolução de problemas ..	11		
Criação de log	11		

Capítulo 1. Visão Geral (Overview)

O IBM® SPSS Analytic Server é uma solução para analítica de Big Data que combina a tecnologia IBM SPSS com sistemas Big Data e permite trabalhar com interfaces com o usuário do IBM SPSS familiares para resolver problemas em uma escala previamente inalcançável.

Porque a Analítica de Big Data É Importante

Os volumes de dados coletados pelas organizações estão crescendo exponencialmente; por exemplo, negócios financeiros e varejistas têm todas as transações dos clientes por um ano (ou dois anos, ou dez anos), provedores telco têm Call Data Records (CDR) e leituras de sensor de dispositivo e as empresas de Internet têm os resultados de crawls da web.

A analítica de big data é necessária onde existe:

- Um grande volume de dados (terabytes, petabytes, exabytes), principalmente quando há uma mistura de dados estruturados & dados não estruturados
- Dados que são alterados/acumulados rapidamente

A analítica de big data também ajuda quando:

- Um grande número (milhares) de modelos está sendo construído
- Modelos são construídos/atualizados frequentemente

Desafios

As mesmas organizações que coletam grandes volumes de dados costumam ter dificuldades ao fazerem uso deles por vários motivos:

- A arquitetura dos produtos de analítica tradicionais não é adequada para cálculo distribuído, e
- Algoritmos estatísticos existentes não são projetados para trabalharem com big data (esses algoritmos esperam que os dados cheguem até a eles, mas os big data são muito caros para se mover); dessa forma
- Executar uma analítica moderna em big data requer novas qualificações e profundo conhecimento de sistemas de big data. Poucos analistas têm essas qualificações.
- As soluções em memória trabalham para solucionar problemas de médio porte, mas não se adaptam muito bem a dados realmente grandes.

Solução

Analytic Server fornece:

- Uma arquitetura centrada em dados que alavanca sistemas de big data, como Hadoop Map/Reduce, como dados em HDFS.
- Uma interface definida para incorporar novos algoritmos estatísticos projetados para acessar dados.
- Interfaces com o usuário do IBM SPSS familiares que ocultam os detalhes dos ambientes de big data para que os analistas possam se concentrar na análise dos dados.
- Uma solução escalável para problemas de qualquer porte.

Arquitetura

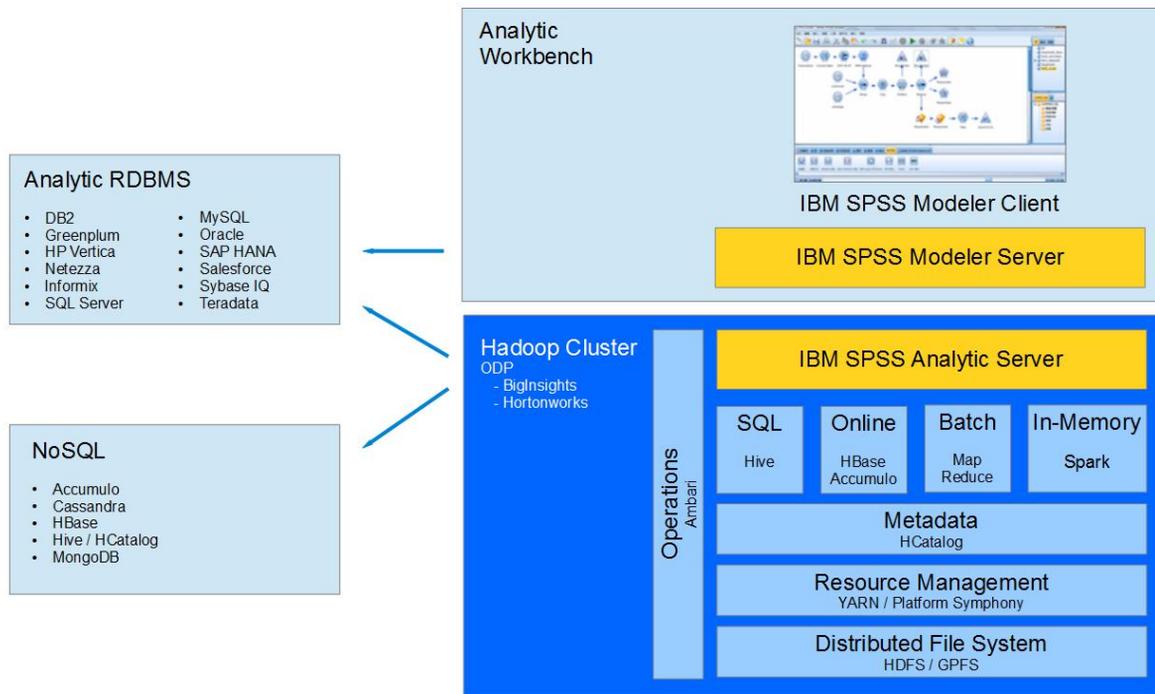


Figura 1. Arquitetura

O Analytic Server está situado entre um aplicativo cliente e uma nuvem Hadoop. Supondo que os dados residem na nuvem, a estrutura geral de tópicos para trabalhar com o Analytic Server está para:

1. Definir as origens de dados do Analytic Server sobre os dados na nuvem.
2. Definir a análise que você deseja executar no aplicativo cliente. Para a liberação atual, o aplicativo cliente é IBM SPSS Modeler.
3. Quando você executa a análise, o aplicativo cliente envia uma solicitação de execução do Analytic Server.
4. O Analytic Server orquestra a tarefa para ser executada na nuvem Hadoop e relata os resultados ao aplicativo cliente.
5. É possível usar os resultados para definir outras análises e as repetições de ciclo.

O Que Há de Novo para Administradores na Versão 2.1

Console do Analytic Server

Função de leitor

Dentro de um locatário, é possível designar usuários e grupos para uma função de Leitor que não pode efetuar login no Analytic Server, mas pode ler as origens de dados do Analytic Server por meio do nó de Origem do Analytic Server no Modeler.

Capítulo 2. Gerenciamento de Arrendatário

Os locatários fornecem uma divisão de alto nível de usuários, projetos e origens de dados para que os objetos não possam ser compartilhados entre locatários. Cada usuário acessa o sistema no contexto de um arrendatário para o qual ele é designado.

Você gerencia locatários e designa usuários a locatários, no console do Analytic Server. A visualização da página Locatários depende da função do usuário que está registrada no console:

- O administrador "super usuário" que é configurado durante a instalação é o gerenciador de locatários. Apenas esse usuário pode criar novos arrendatários e editar as propriedades de qualquer arrendatário.
- Os usuários com função de Administrador podem editar as propriedades do arrendatário ao qual estão conectados.
- Os usuários com função de Usuário não podem editar propriedades do arrendatário. A página Locatários é ocultada deles.
- Usuários com a função de Leitor não podem editar origens de dados ou mesmo efetuar login no console do Analytic Server.

Os administradores podem acessar as páginas Projetos e Origens de Dados e gerenciar qualquer projeto ou origem de dados para limpeza e administração. Consulte o *IBM SPSS Analytic Server Guia do Usuário* para obter mais informações.

Lista de locatários

A página principal Locatários exibe os locatários existentes em uma tabela. Apenas o administrador "super usuário" pode fazer edições nessa página.

- Clique no nome de um locatário para exibir seus detalhes e editar suas propriedades.
- Clique na URL de um locatário para abrir o console no contexto desse locatário.

Nota: Você será desconectado do console e precisará efetuar login com credenciais válidas para o locatário.

- Digite na área de procura para filtrar a lista a fim de exibir apenas os locatários com a sequência de caracteres de procura em seu nome.
- Clique em **Novo** para criar um novo locatário com o nome especificado no diálogo **Incluir novo locatário**. Consulte "Regras de nomenclatura" na página 4 para restrições nos nomes que você pode atribuir aos locatários.
- Clique em **Excluir** para remover o(s) locatário(s) selecionado(s).
- Clique em **Atualizar** para atualizar a lista.

Detalhes de locatário individual

A área de conteúdo é dividida em várias seções redutíveis.

Detalhes

Nome Um campo de texto editável que exibe o nome do arrendatário.

Descrição

Um campo de texto editável que permite fornecer texto explicativo sobre o arrendatário.

URL Esta é a URL a ser fornecida aos usuários para efetuarem login no arrendatário através do

console do Analytic Server, e a ser usada para configurar o servidor SPSS Modeler. Consulte o Guia de Instalação e Configuração *IBM SPSS Analytic Server* para obter detalhes sobre configurar o SPSS Modeler.

Status Ativo os locatários estão atualmente em uso. Tornar um locatário **Inativo** evita que os usuários efetuem login para esse locatário, mas não exclui nenhuma informação subjacente.

Diretores

Os diretores são usuários e grupos extraídos do provedor de segurança que é configurado durante a instalação. É possível incluir diretores em um arrendatário como Administradores ou Usuários.

- Digitando os filtros da caixa de texto nos usuários e grupos com a sequência de caracteres de procura em seu nome. Selecione **Administrador**, **Usuário** ou **Leitor** na lista suspensa para designar suas funções dentro do locatário. Clique em **Incluir participante** para incluí-lo na lista de autores.
- Para remover um participante, selecione um usuário ou grupo na lista de membros e clique em **Remover participante**.

Métricas

Permite configurar limites de recurso para um locatário. Relata o espaço em disco atualmente usado pelo locatário.

- É possível configurar uma cota de espaço em disco máxima para o locatário; quando este limite é atingido, mais nenhum dado pode ser gravado no disco neste locatário até que seja liberado espaço em disco suficiente para colocar o uso do espaço em disco do locatário abaixo da cota.
- É possível configurar um nível de aviso de espaço em disco para o locatário; quando a cota é excedida, nenhuma tarefa analítica pode ser enviada por diretores neste locatário até que seja liberado espaço em disco suficiente para colocar o uso do espaço em disco do locatário abaixo da cota.
- É possível configurar um número máximo de tarefas paralelas que podem ser executadas uma única vez neste locatário; quando a cota é excedida, nenhuma tarefa analítica pode ser enviada por diretores neste locatário até que uma tarefa em execução atualmente seja concluída.
- É possível configurar o número máximo de campos que uma origem de dados pode ter. O limite é verificado sempre que uma origem de dados é criada ou atualizada.
- É possível configurar o tamanho máximo do arquivo em megabytes. O limite é verificado quando um arquivo é atualizado.

Configuração do provedor de segurança

Permite especificar o provedor de autenticação do usuário. **Padrão** usa o provedor do locatário padrão, que foi configurado durante a instalação e a configuração. **LDAP** permite autenticar usuários com um servidor LDAP externo, como Active Directory ou OpenLDAP. Especifique as configurações para o provedor e opcionalmente as configurações de filtro para controlar os usuários e grupos disponíveis na seção Diretores.

Regras de nomenclatura

Para qualquer coisa que possa receber um nome exclusivo no Analytic Server, como origens de dados e projetos, as regras a seguir são aplicadas a esses nomes.

- Em um único locatário, os nomes devem ser exclusivos dentro de objetos do mesmo tipo. Por exemplo, duas origens de dados não podem ser denominadas insuranceClaims, mas uma origem de dados e um projeto poderiam cada um ser denominados insuranceClaims.
- Os nomes fazem distinção entre maiúsculas e minúsculas. Por exemplo, insuranceClaims e InsuranceClaims são considerados nomes exclusivos.
- Os nomes ignoram espaço em branco à esquerda e à direita.
- Os caracteres a seguir são inválidos nos nomes.

~, #, %, &, *, {, }, \\, :, <, >, ?, /, |, ", \t, \r, \n

Capítulo 3. Introdução para usuários

Informe os usuários para que naveguem até `http://<host>:<port>/<context-root>/admin/<tenant>` e insiram seu nome de usuário e senha para efetuar logon no console do Analytic Server.

<host>

O endereço do host do Analytic Server.

<port>

A porta pela que Analytic Server está recebendo. Por padrão, é 8080.

<context-root>

A raiz de contexto do Analytic Server. Por padrão, é `analyticserver`.

<tenant>

Em um ambiente de diversos locatários, o locatário ao qual você pertence. Em um ambiente de único locatário, o locatário padrão é **ibm**.

Por exemplo, se a máquina host tem o endereço IP 9.86.44.232, se foi criado o locatário "mycompany" e usuários foram incluídos nele e se outras configurações foram mantidas no padrão, então, os usuários terão de navegar até o `http://9.86.44.232:8080/analyticserver/admin/mycompany` para acessar o console do Analytic Server.

Capítulo 4. Nomes de tarefa do Analytic Server

O Analytic Server produz tarefas de redução de mapa, que podem ser monitoradas por meio de seu rastreador de tarefas do cluster Hadoop.

O nome da tarefa de redução de mapa possui a estrutura a seguir.

AS/{tenant name}/{user name}/{algorithm name}

{tenant name}

Este é o nome do locatário sob o qual a tarefa é executada.

{user name}

Este é o usuário que solicitou a tarefa.

{algorithm name}

Este é o algoritmo primário na tarefa. Observe que um único fluxo pode gerar diversas tarefas de redução de mapa; da mesma forma, várias operações em um fluxo podem estar contidas em uma única tarefa de redução de mapa.

Capítulo 5. Resolução de problemas

O Analytic Server fornece várias ferramentas úteis para a determinação de problema.

Criação de log

O Analytic Server cria arquivos de log do cliente e arquivos de rastreamento que são úteis para diagnosticar problemas. Com a instalação padrão do Liberty, é possível localizar os arquivos de log no diretório {AS_ROOT}/ae_wlpsrver/usr/servers/aeserver/logs.

A configuração padrão da criação de log produz dois arquivos de log que são substituídos em uma base diária.

as.log Este arquivo contém o resumo de alto nível de mensagens de aviso e erro informativas. Verifique esse arquivo primeiro quando ocorrerem erros do servidor que não podem ser resolvidos usando a mensagem de erro que é exibida na Interface com o usuário.

as_trace.log

Este arquivo contém todas as entradas de as.log, mas inclui mais informações que são primariamente voltadas ao suporte IBM e ao desenvolvimento para fins de depuração.

Analytic Server usa Apache LOG4J como seu recurso de criação de log subjacente. Usando LOG4J, a criação de log pode ser dinamicamente ajustada editando o arquivo de configuração {AS_SERVER_ROOT}/configuration/log4j.xml. O Suporte pode solicitar que faça isso para ajudar a diagnosticar problemas ou você pode querer modificar isto para limitar o número de arquivos de log mantidos. Mudanças no arquivo são automaticamente detectadas em poucos segundos, assim o Analytic Server não precisa ser reiniciado.

Para obter mais informações sobre log4j e o arquivo de configuração, consulte a documentação no website Apache oficial em <http://logging.apache.org/log4j/>.

Informações de versão

É possível determinar que versão do Analytic Server é instalada verificando a pasta {AS_ROOT}/properties/version. Os arquivos a seguir contêm informações da versão.

IBM_SPSS_Analytic_Server-*.swtag

Contém informações detalhadas do produto.

version.txt

Versão e número da construção para o produto instalado.

Coletor de logs

Quando os problemas não podem ser resolvidos diretamente revisando os arquivos de log, é possível compactar todos os logs e enviá-los ao suporte IBM. Há um utilitário que é fornecido para tornar mais simples a coleta de todos os dados necessários.

Usando um shell de comando, execute os comandos a seguir:

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

Esses comandos criam um arquivo compactado em {AS_ROOT}/bin. O arquivo compactado contém todos os arquivos de log e informações de versão do produto.

Problemas Comuns

Esta seção descreve alguns problemas de administração comuns e como é possível corrigi-los.

Segurança

A autenticação do Kerberos falhará na tentativa de acessar a origem de dados HCatalog

Se forem detectados erros como os seguintes:

cause:javax.security.sasl.SaslException: Failure to initialize security context

com.spss.analyticframework.api.exceptions.ComponentException: Cannot access HCatalog

Será necessário garantir que o TGT do Kerberos do usuário HDFS esteja em cache e disponível no host do servidor do Analytic Server. Para executar isso:

1. Pare o processo do Analytic Server.
2. Execute `kinit -f $hdfs.user` no host do Analytic Server, onde **\$hdfs.user** é definido no arquivo `config.properties` e possui permissão de gravação para a raiz analítica
3. Inicie o Analytic Server.

Console do Analytic Server

Acessando o Console do Analytic Server no Safari em iOS

A lista suspensa de status do locatário não funciona e não é possível atualizar o modelo de dados para origens de dados baseadas em Arquivos. Use outro navegador ao executar estas ações.

Fluxos em execução

A execução das tarefas PySpark falhou

1. No Console Ambari, inclua `export SPARK_HOME=/usr/iop/current/spark-client` para o parâmetro **yarn-env** do serviço YARN.
2. Reinicie o serviço YARN.

Avisos

Estas informações foram desenvolvidas para produtos e serviços oferecidos nos Estados Unidos.

É possível que a IBM não ofereça os produtos, serviços ou recursos discutidos neste documento em outros países. Consulte seu representante IBM local para obter informações sobre os produtos e serviços disponíveis atualmente em sua área. Qualquer referência a um produto, programa ou serviço IBM não significa que apenas produtos, programas ou serviços IBM possam ser usados. Qualquer produto, programa ou serviço funcionalmente equivalente, que não infrinja nenhum direito de propriedade intelectual da IBM poderá ser usado em substituição. Entretanto, a avaliação e verificação da operação de qualquer produto, programa ou serviço não IBM são de responsabilidade do usuário.

A IBM pode ter patentes ou solicitações de patentes pendentes relativos a assuntos tratados nesta publicação. O fornecimento desta publicação não lhe garante direito algum sobre tais patentes. Pedidos de licença devem ser enviados, por escrito, para:

Gerência de
Relações Industriais e Comerciais da IBM
Brasil
Av. Pasteur, 138-146, Botafogo
Rio de Janeiro, RJ
CEP 22290-240.

Para pedidos de licença relacionados a informações de DBCS (Conjunto de Caracteres de Byte Duplo), entre em contato com o Departamento de Propriedade Intelectual da IBM em seu país ou envie pedidos de licença, por escrito, para:

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

O parágrafo a seguir não se aplica ao Reino Unido ou qualquer outro país em que tais disposições não estejam de acordo com a legislação local: A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO "NO ESTADO EM QUE SE ENCONTRA", SEM GARANTIA DE NENHUM TIPO, SEJA EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS A ELAS NÃO SE LIMITANDO, AS GARANTIAS IMPLÍCITAS DE NÃO INFRAÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO PROPÓSITO. Alguns países não permitem a exclusão de garantias expressas ou implícitas em certas transações; portanto, essa disposição pode não se aplicar ao Cliente.

Essas informações podem conter imprecisões técnicas ou erros tipográficos. São feitas alterações periódicas nas informações aqui contidas; tais alterações serão incorporadas em futuras edições desta publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar os produtos e/ou programas descritos nesta publicação, sem aviso prévio.

Referências nestas informações a websites não IBM são fornecidas apenas por conveniência e não representam de forma alguma um endosso a esses websites. Os materiais contidos nesses websites não fazem parte dos materiais deste produto IBM e a utilização desses websites é de inteira responsabilidade do Cliente.

A IBM pode usar ou distribuir as informações fornecidas da forma que julgar apropriada sem incorrer em qualquer obrigação para com o Cliente.

Licenciados deste programa que desejam obter informações sobre esse assunto com objetivo de permitir: (i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) a utilização mútua das informações trocadas, devem entrar em contato com:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

Tais informações podem estar disponíveis, sujeitas a termos e condições apropriados, incluindo em alguns casos o pagamento de uma taxa.

O programa licenciado descrito neste documento e todo o material licenciado disponível são fornecidos pela IBM sob os termos do IBM Customer Agreement, Contrato de Licença do Programa Internacional da IBM ou qualquer contrato equivalente.

Todos os dados de desempenho aqui contidos foram determinados em um ambiente controlado. Portanto, os resultados obtidos em outros ambientes operacionais podem variar significativamente. Algumas medidas podem ter sido tomadas em sistemas em nível de desenvolvimento e não há garantia de que estas medidas serão as mesmas nos sistemas geralmente disponíveis. Além disso, algumas medidas podem ter sido estimadas por meio de extrapolação. Os resultados reais podem variar. Os usuários deste documento devem verificar os dados aplicáveis aos seus ambientes específicos.

As informações relativas a produtos não IBM foram obtidas junto aos fornecedores dos respectivos produtos, de seus anúncios publicados e de outras fontes disponíveis atualmente. A IBM não testou estes produtos e não pode confirmar a precisão do desempenho, da compatibilidade ou de qualquer outra reivindicação relacionada a produtos não IBM. Dúvidas sobre os recursos de produtos não IBM devem ser encaminhadas diretamente a seus fornecedores.

Todas as instruções relativas à orientação ou intenção futura da IBM estão sujeitas a alterações ou retiradas sem aviso prévio e representam apenas metas e objetivos.

Todos os preços IBM mostrados são preços de varejo sugeridos pela IBM, são atualizados e estão sujeitos a alterações sem aviso prévio. Os preços dos revendedores podem variar.

Estas informações são apenas para fins de planejamento. As informações nesta publicação estão sujeitas a alterações antes que os produtos descritos se tornem disponíveis.

Estas informações contêm exemplos de dados e relatórios utilizados em operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos podem incluir nomes de indivíduos, empresas, marcas e produtos. Todos esses nomes são fictícios e qualquer semelhança com nomes e endereços utilizados por uma empresa real é mera coincidência.

Cada cópia ou parte desses programas de amostra ou qualquer trabalho derivado deve incluir um aviso de copyright com os dizeres:

Estas informações contêm exemplos de dados e relatórios utilizados em operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos podem incluir nomes de indivíduos, empresas, marcas e produtos. Todos esses nomes são fictícios e qualquer semelhança com nomes e endereços utilizados por uma empresa real é mera coincidência.

Cada cópia ou parte desses programas de amostra ou qualquer trabalho derivado deve incluir um aviso de copyright com os dizeres:

© nome da sua empresa) (ano). Partes deste código são derivadas dos Programas de Amostra da IBM Corp.

© Copyright IBM Corp. _digite o ano ou anos_. Todos os direitos reservados.

Se estiver visualizando essas informações em formato eletrônico, as fotografias e ilustrações coloridas podem não aparecer.

Marcas Registradas

IBM, o logotipo IBM e ibm.com são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em vários países no mundo todo. Outros nomes de empresas, produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual de marcas comerciais IBM está disponível na web em "Copyright and trademark information" em www.ibm.com/legal/copytrade.shtml.

Adobe, o logotipo Adobe, PostScript e o logotipo PostScript são marcas comerciais ou marcas registradas da Adobe Systems Incorporated nos Estados Unidos e/ou em outros países.

IT Infrastructure Library é uma marca registrada da Central Computer and Telecommunications Agency que agora faz parte do Departamento de Comércio do Governo.

Intel, logotipo Intel, Intel Inside, logotipo Intel Inside, Intel Centrino, logotipo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium são marcas comerciais ou marcas registradas da Intel Corporation ou suas subsidiárias nos Estados Unidos e em outros países.

Linux é uma marca registrada da Linus Torvalds nos Estados Unidos e/ou em outros países.

Microsoft, Windows, Windows NT e o logotipo Windows são marcas registradas da Microsoft Corporation nos Estados Unidos e/ou em outros países.

ITIL é uma marca registrada e uma marca de comunidade registrada da The Minister for the Cabinet Office e está registrada no U.S. Patent and Trademark Office.

UNIX é uma marca registrada da The Open Group nos Estados Unidos e em outros países.

Cell Broadband Engine é uma marca comercial da Sony Computer Entertainment, Inc. nos Estados Unidos e/ou em outros países e é utilizada sob licença.

Linear Tape-Open, LTO, o logotipo LTO, Ultrium e o logotipo Ultrium são marcas comerciais da HP, IBM Corp. e Quantum nos Estados Unidos e em outros países.



Impresso no Brasil