

Usando o Entity Analytics para Aumentar Muito a Precisão de Seus Modelos Mais Rápida e Facilmente



Redguides

para Líderes de Negócios

Dra. Lisa Sokol
Jeff Jonas

- Aprenda como o Entity Analytics fornece valor para seus negócios
- Veja como o IBM SPSS Modeler Premium suporta o Entity Analytics
- Obtenha insight no Entity Analytics em tempo real



Visão Geral do Entity Analytics

Analistas rotineiramente enfrentam desafios excessivos conforme tentam integrar diferentes dados de toda empresa. Esta afirmação é especialmente verdadeira quando estes dados contêm variabilidade natural (por exemplo, Bob versus Robert), erros não intencionais (como um mês e um dia transpostos em um data de nascimento) e mentiras fabricadas profissionalmente (como uma identidade falsa). A integração incorreta ou incompleta pode afetar negativamente qualquer solução analítica que seja construída usando os dados.

Ao implementar o Entity Analytics, os analistas podem superar alguns dos desafios mais difíceis de preparação de dados com facilidade sem precedentes. Ao usar o Entity Analytics, os analistas podem gerar maior qualidade, modelos analíticos mais precisos que resultam em melhores resultados de negócios. Esta atividade pode ser realizada independentemente de o objetivo estar detectando e priorizando risco ou reconhecendo e respondendo a uma oportunidade.

Uma atividade de preparação de dados críticos envolve reconhecer quando diversas referências à mesma entidade são a mesma entidade (na mesma e através de origens de dados). Por exemplo, é essencial entender a diferença entre três transações executadas por três pessoas diferentes versus uma pessoa que executou todas as três transações.

Dada a determinação quando entidades são as mesmas (resolvidas), um entendimento ainda mais profundo é alcançado ao reconhecer quando estas entidades resolvidas são relacionadas entre si (como compartilhando um endereço inicial). Indo além da correspondência simplista ou de tecnologias de mesclagem do passado, o Entity Analytics entrega algo novo: *acumulação de contexto* verdadeira. A acumulação de contexto é o processo incremental de relacionar novos dados a dados antigos e lembrar destes relacionamentos. É possível entender algo melhor levando-se em conta as informações ao seu redor. Este processo resulta na precisão melhorada de dados.

Por exemplo, uma peça de quebra-cabeça independente pode ser difícil de avaliar para importância quando você a observa sozinha (mostrado na Figura 1 na página 2). Entretanto, ao comparar primeiro a peça com todo o quebra-cabeça, para ver como ela se relaciona com as peças anteriores dele, é possível entender melhor a imagem como um todo e fazer uma predição melhor.

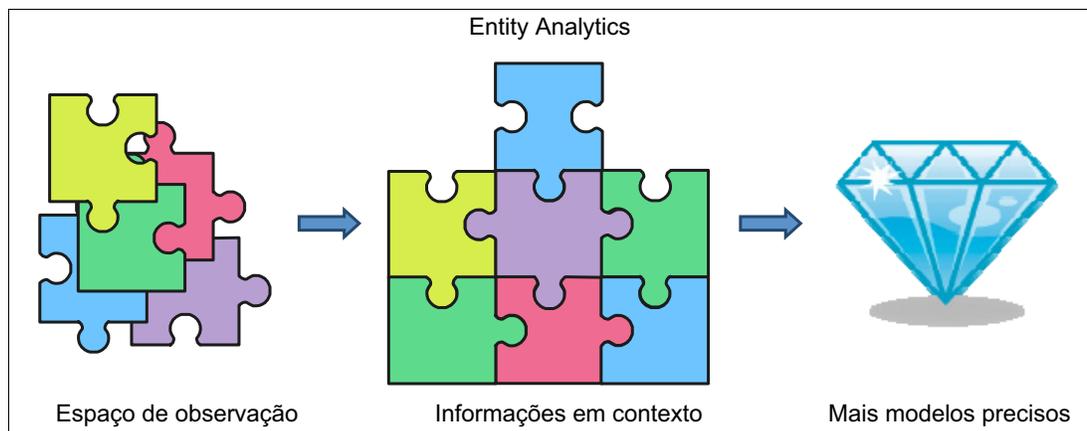


Figura 1 Entity Analytics acumulando contexto sobre diferentes dados

Esta publicação IBM® Redguide™ explica como o Entity Analytics pode ajudar os analistas através dos modelos que eles criam para orientar melhores resultados de negócios. Este guia fornece uma visão geral do produto IBM SPSS® Modeler Premium que incorpora os recursos do Entity Analytics. Ele também fornece exemplos de como o Entity Analytics faz determinações e ajuda a resolver problemas de negócios. Além disso, este guia o leva através de um cenário condutor para um banco fictício para ilustrar o uso do Entity Analytics.

Sobre o IBM SPSS Modeler Premium

O IBM SPSS Modeler Premium é um ambiente de trabalho preditivo e de análise de texto de alto desempenho que ajuda a obter insight sem precedentes de seus dados. Ele fornece um amplo conjunto de recursos analíticos, incluindo os recursos a seguir:

- ▶ Visualização e exploração de dados
- ▶ Manipulação de dados
- ▶ Limpeza e transformação de dados
- ▶ Criação e avaliação de modelos preditivos
- ▶ Implementação de resultados na forma de modelos de produção (tempo de execução) ou pontuações

Funcionalidade do Entity Analytics no IBM SPSS Modeler Premium

O SPSS Modeler Premium contém os recursos do Entity Analytics que os analistas podem usar para associar rapidamente dados de identidade, comportamento e ação com suas respectivas identidades em tempo real ou em lote, com facilidade extraordinária. Estes recursos do Entity Analytics no SPSS Modeler Premium representam uma tecnologia avançada, a primeira de seu tipo que está comercialmente disponível. Melhor ainda, estes recursos são fáceis de usar, de forma que seja possível iniciar imediatamente se beneficiar dele.

Historicamente, os analistas gastavam até 80% de seu tempo preparando e limpando dados para análise. Usando o Entity Analytics, os usuários agora podem construir modelos de dados muito mais precisos, com base em dados de limpador em um curto prazo.

Os usuários do Entity Analytics obtêm as vantagens distintas a seguir:

- ▶ Mais figura precisa
Quanto mais identificadores que acumulam para uma entidade, mais precisa a tecnologia do Entity Analytics se torna.
- ▶ Melhores modelos
Informações em contexto (entendendo como os dados se relacionam) entregam modelos de qualidade mais alta.
- ▶ Melhores resultados
Modelos de qualidade mais altas aplicados a transações de contexto aprimorado produzem melhores decisões (por exemplo, cálculos de pontuação de risco).

Por exemplo, uma prática de regulamentações comum é requerer dos bancos que eles relatem todas as transações acima de US\$5.000. Para isto, os bancos devem ser capazes de entender a diferença entre cinco transações de depósito de dinheiro de US\$1.000 aparentemente não relacionadas versus uma pessoa que transaciona um depósito de dinheiro de US\$5.000. Se um banco não puder qualificar com precisão as transações cumulativas (históricas) para este indivíduo, ele não é capaz de determinar se o limite de US\$5.000 foi cruzado.

O Entity Analytics (mostrado na Figura 2) fornece um meio fácil (usando acumulação de contexto) para associar as transações para a entidade correta, apesar da falta de chave comum. (As contas não compartilham um número de ID de imposto.) Como um resultado, quando as transações estão em contexto, os modelos de pontuação operam no número de US\$5.000, não um número aparentemente não relacionado de transações de US\$1.000.

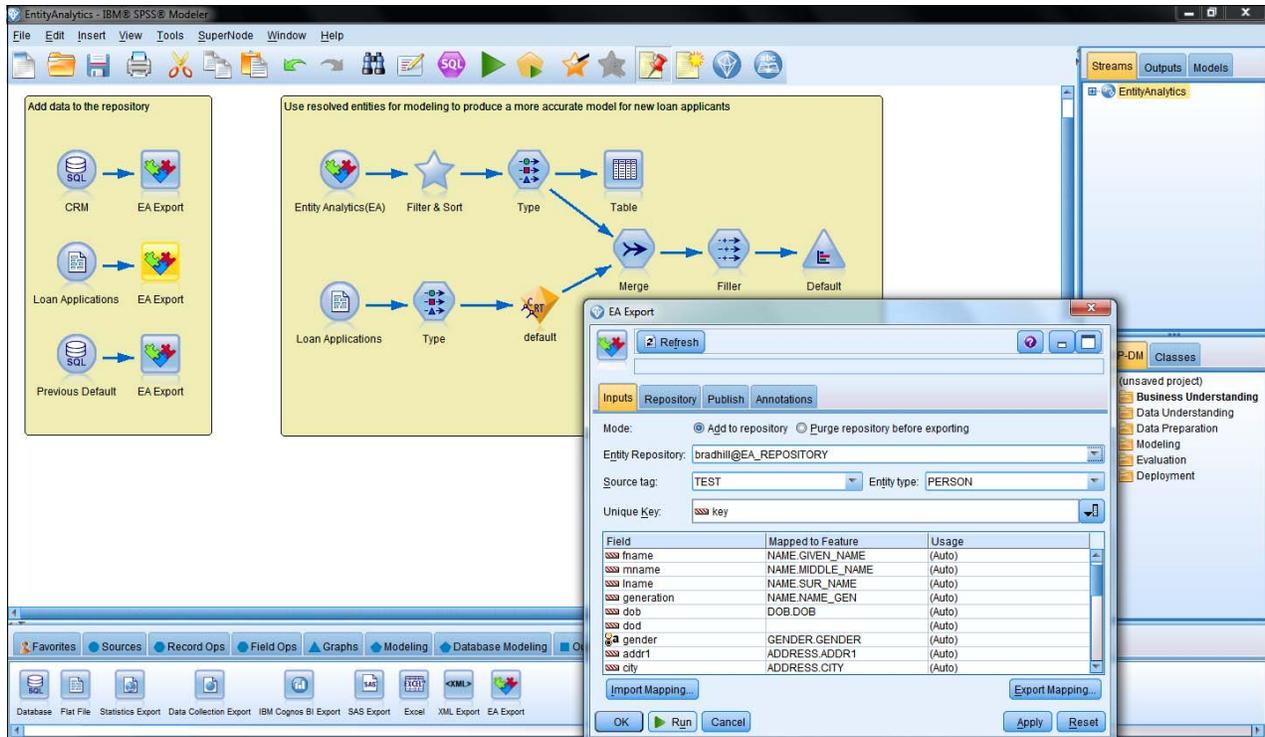


Figura 2 Exemplo de um fluxo de SPSS Modeler Premium que usa Entity Analytics

No SPSS Modeler Premium, o Entity Analytics pode ser usado das maneiras a seguir:

- ▶ O nó de exportação do Entity Analytics executa a acumulação de contexto. Este nó determina se duas entidades (como indivíduos, corporações ou veículos) são a mesma. Esta atividade é realizada apesar do fato de que elas foram registradas separadamente e, em um grau, registradas diferentemente. Se as entidades são determinadas como sendo a mesma, seus identificadores (como nome, endereço ou telefone) e suas medições (como balanço médio ou limite de crédito) são acumulados para esta entidade. Este nó aplica automaticamente técnicas de correspondência difusa sofisticadas. Por exemplo, ela se aproveita de uma biblioteca interna com base em mais de 800 milhões de nomes de pessoas para entregar uma comparação de nomes totalmente culturalmente conscientes em todo o mundo. Conforme as entidades são resolvidas, o entendimento sobre cada entidade melhora. O nó de exportação é usado frequentemente para integrar dados históricos com novos dados incrementais.
- ▶ Ao usar o nó de origem do Entity Analytics (que lê identidades resolvidas) um analista pode acessar as informações *no contexto*. Este nó é frequentemente usado para analisar informações históricas (em contexto) e ao criar visualizações de dados para suportar novas construções de modelos.
- ▶ O nó do Entity Analytics de fluxo é usado para aplicar novos registros (em lote ou tempo real) nas informações históricas. Ele instantaneamente reconhece quando as entidades são a mesma ou estão relacionadas. Esta capacidade é mais ou menos parecida a dar uma peça de dados (uma peça de quebra-cabeça) e perguntar qual outras peças relacionadas existem (as peças de quebra-cabeça associadas). Dado que este nó descobre novos dados sobre uma entidade, o novo conhecimento pode ser passado para um processo que racionaliza os dados sobre uma entidade.

Por exemplo, os novos dados recém-descobertos que você tem em outra conta bancária podem ser usados para atualizar uma variável de riqueza. Processos de avaliação de padrão de recebimento de dados podem analisar esta entidade novamente com padrões e modelos de interesse, para determinar se ela agora satisfaz um limite interessante. O padrão ou análise de pontuação será mais preciso, graças ao contexto aprimorado.

Cenário condutor do banco que demonstra o uso do Entity Analytics

Para ver como o Entity Analytics funciona, considere este exemplo hipotético que envolve um processo bancário típico de empréstimos feitos a clientes.

A análise preditiva pode ser usada para ajudar um banco a determinar quais clientes provavelmente pagarão seus empréstimos versus a inadimplência em seus empréstimos. Para determinar a probabilidade de um indivíduo pagar um empréstimo, modelos são criados usando dados disponíveis de várias origens como os exemplos a seguir:

- ▶ Dados do cliente histórico (como receita, débito ou inadimplências anteriores)
- ▶ Principais resultados passados (como limite de crédito, quantia de pagamento média e delinquências)
- ▶ Outros pontos de dados usados frequentemente

A Figura 3 mostra um exemplo de dados para empréstimo do cliente do banco. As primeiras duas linhas são dados históricos do cliente. A terceira linha contém dados para um cliente que está se candidatando a um novo empréstimo. A última coluna na tabela indica se o cliente possui uma aplicativo de empréstimos pendente.

Cliente Número	Receita	Crédito Dívida	Outro Dívida	Dívida Para Receita	Anterior Padrão	Pendente
102	8000	5359	2009	92.1	S	N
343	9000	6000	3000	100	S	N
642	31000	1362	4001	17.3	N	S

Figura 3 Dados principais do banco

Ao usar dados históricos, o SPSS Modeler Premium pode gerar um modelo preditivo que pode analisar novas solicitações de empréstimo para a probabilidade de reembolso. Um exemplo de uma regra de pontuação gerada pode ser: “Se um indivíduo possui um débito para proporção de receita maior que 24,6 e inadimplências anteriores, eles provavelmente não pagarão um empréstimo futuro.” No exemplo, mostrado em Figura 3, a entidade #642 está se candidatando a um empréstimo. Esta pessoa afirma não ter nenhuma inadimplência anterior e ter uma proporção de débito para receita baixa. Ao usar a regra definida anteriormente como um avaliador, o indivíduo pode ser aprovado para receber um empréstimo.

Se você olhar mais atentamente na Figura 3 na página 5, é possível imaginar a diferença entre os três pontos de dados sobre três clientes diferentes versus estes três pontos de dados sobre o mesmo cliente. Suponha que o cliente #642 seja a mesma pessoa que #102 e #343. Você consideraria que este cliente (que possui a requisição de crédito pendente) seria um risco de crédito se soubesse com alguma confiança que esta pessoa foi inadimplente duas vezes no passado?

Se os clientes usaram seus nomes, endereços e identificadores verdadeiros consistentemente e forneceram todos os detalhes de forma abrangente e não ambígua, determinar se estas informações são sobre o mesmo cliente poderia ser trivial. Infelizmente, graças a problemas de qualidade de dados não intencionais e intenção criminal periódica, determinar se estas informações representam o mesmo cliente é mais fácil falar do que fazer. Felizmente, com o Entity Analytics, os usuários podem rápida e facilmente executar acumulação de contexto para detectar exatamente esta situação e mais.

A Figura 4 mostra que as entidades #102, #343 e #642 compartilham identificadores suficientes e fazem uma forte afirmação de que estas entidades são o mesmo cliente.

Entity 102	Entity 343	Entity 642	Resolved Entity
Name Beth L. Doe-Smith BL Doe	Full Liz Doe	Full Elizabeth Lisa Doe	Name Elizabeth Lisa Doe Liz Doe Beth L Doe-Smith BL Doe
Addr1 123 Main Street 777 Park Road	Addr1 33 Red Dr	Addr1 33 Reed Dr	Addr1 123 Main Street 777 Park Road 33 Red Dr 33 Reed Dr
City New York	City Mamaroneck	City White Plains	City New York, White Plains, Mamaroneck
State NY	State NY	State NY	State NY
Phone 9587331234	Postal 10354	Postal 10354	Postal 11732, 10354
DOB 6/21/1954	Phone 958-733-1234 959-698-2234	Phone 959-698-2234	Phone 958-733-1234 959-698-2234
Income \$8,000	Income \$9,000	Income \$31,000	DOB 6/21/1954
Credit Debt \$5,359	Credit Debt \$6,000	Credit Debt \$1,362	Defaults Yes
Other Debt \$2,009	Other Debt \$3,000	Other Debt \$4,001	Income \$48,000
Debt to Income 92.1	Debt to Income 100	Debt to Income 17.3	Credit Debt \$12,722
Prev Default? True	Prev Default? True	Prev Default? False	Other Debt \$9,009
Pending Loan False	Pending Loan False	Pending Loan True	Debt to Income 113.5
			Prev Default? True
			Pending Loan True

Figura 4 Atributos comuns através de diferentes recursos que são usados para construir contexto

O uso destes fatos coletados na coluna *Entidade Resolvida* destaca o contexto essencial para ajudar a pontuar adequadamente a requisição de empréstimo pendente para a entidade #642. Ao usar os dados da entidade criados pelo nó de origem do Entity Analytics, um analista pode somar o débito de crédito verdadeiro. O analista pode determinar que a entidade resolvida possui um débito de crédito de US\$12.722 e possui uma proporção de débito para receita de 113,5. Quando o algoritmo de pontuação é aplicado à entidade resolvida, a pontuação indica que a entidade #642 não deve receber o empréstimo. Este exemplo demonstra o valor verdadeiro do Entity Analytics, que é proporcionar decisões mais precisas, mais rapidamente.

Entity Analytics em tempo real

Ao usar o Entity Analytics no SPSS Modeler Premium, as empresas podem analisar transações em tempo real para tomar decisões ótimas em contexto. Com base em todas as informações “totais”, os modelos podem prever resultados mais precisamente para tomada de decisão instantânea, como detecção de fraude em tempo real.

Imagine um investigador de fraudes que acabou de encontrar um novo endereço que esteja relacionado a uma investigação de crime interna em andamento. Com estas informações, alguns segundos depois o Entity Analytics alerta este investigador de que um funcionário no departamento de crédito do investigador possui o mesmo endereço. Através deste processo de acumulação de contexto, o Entity Analytics relacionou os novos dados (novo endereço) a dados anteriores (investigação, clientes e funcionários), entregando este extraordinário insight de *ameaça interna* e muito mais.

Resumo

Ao usar o recurso do Entity Analytics no IBM SPSS Modeler Premium, os analistas podem puxar diferentes dados corporativos juntos no contexto. As organizações podem, então, usar estas informações no contexto para melhorar a qualidade do modelo, tomar melhores

decisões e finalmente atingir sucesso maior, independentemente de o objetivo minimizar riscos ou reconhecer oportunidades.

Uma organização que pode entender o que conhece e fazer algo sobre isso mais rápido que a concorrência é mais competitiva. Com esta nova tecnologia animadora, organizações de todos os tamanhos podem obter este limite competitivo hoje.

O software de analítica de negócios da IBM entrega insights acionáveis que os tomadores de decisão precisam para alcançar melhor desempenho dos negócios. A IBM oferece um portfólio abrangente unificado de business intelligence, analítica preditiva e avançada, gerenciamento de estratégia e desempenho financeiro, controle, risco e conformidade, além de aplicativos de analítica.

Com o software IBM, as empresas podem identificar tendências, padrões e anomalias; comparar cenários *e se*; prever potenciais ameaças e oportunidades; identificar e gerenciar riscos de negócios chave; e planejar, fazer orçamento e prever recursos. Com estes recursos analíticos profundos, nossos clientes em todo o mundo podem entender, priorizar e moldar melhor os resultados de negócios.

Outros recursos para obter mais informações

Para obter mais informações sobre o SPSS Modeler Premium, consulte a página do produto em:

ibm.com/software/analytics/spss/products/modeler/premium.html

A equipe do quem escreveu este guia

Este guia foi produzido por uma equipe de especialistas do mundo todo que está trabalhando com o International Technical Support Organization (ITSO).

Dra. Lisa Sokol é uma arquiteta no escritório do IBM Government Services CTO. Suas áreas primárias de interesse são ajudar comunidades governamentais em lidar com o problema de sobrecarga de decisões e usar analíticas para descobrir informações acionáveis enterradas em grandes quantias de dados. Ela projetou vários sistemas que detectam e analisam risco de ameaças que é relativo a fraude, terrorismo, contra-inteligência e atividade criminal. Ela possui um doutorado em pesquisa de operações da Universidade de Massachusetts.

Jeff Jonas é um IBM Fellow e Chief Scientist do IBM Entity Analytics Group. Antes de se associar à IBM, ele liderou sua empresa, Pesquisa e Desenvolvimento de Sistemas, através do projeto e desenvolvimento de vários sistemas exclusivos. Ele projeta tecnologia de última geração que ajuda as organizações a obter melhores vantagens de seus ativos de informações de toda a empresa. Ele viaja por todo o mundo e fala sobre inovação, segurança nacional e privacidade com líderes governamentais, executivos do segmento de mercado, laboratórios de ideias globais líderes, grupos de advocacia de privacidade e organizações de pesquisa política. Ele é um membro da Markle Foundation Task Force on National Security in the Information Age e um membro do conselho do US Geospatial Intelligence Foundation (USGIF), do EPIC Advisory Board e do Privacy International Advisory Board. Ele também é um Associado Sênior no Center for Strategic and International Studies (CSIS) e um Distinguished Engineer of Information Systems (auxiliar) na Singapore Management University.

Obrigado a estas pessoas por suas contribuições a este projeto:

LindaMay Patterson
ITSO, Rochester, MN

Agora é possível se tornar um autor publicado também!

Aqui está uma oportunidade para destacar suas qualificações, crescer em sua carreira e se tornar um autor publicado – tudo ao mesmo tempo! Junte-se ao projeto de residência da ITSO e ajude a escrever um livro em sua área de experiência, enquanto afia sua experiência usando tecnologias de ponta. Seus esforços ajudarão a aumentar a aceitação do produto e a satisfação do cliente, conforme você expande sua rede de contatos técnicos e relacionamentos. As residências têm de duas a seis semanas de duração, e é possível participar pessoalmente ou como um residente remoto trabalhando de sua casa.

Encontre mais sobre o programa de residência, navegue pelo índice de residência e se cadastre online em:

ibm.com/redbooks/residencies.html

Fique conectado com o IBM Redbooks

- ▶ Encontre-nos no Facebook:
<http://www.facebook.com/IBMRedbooks>
- ▶ Siga-nos no Twitter:
<http://twitter.com/ibmredbooks>
- ▶ Procure por nós no LinkedIn:
<http://www.linkedin.com/groups?home=&gid=2130806>
- ▶ Explore novas publicações do IBM Redbooks®, residências e workshops com a newsletter semanal do IBM Redbooks:
<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>
- ▶ Fique atualizado sobre publicações recentes do Redbooks com Feeds RSS:
<http://www.redbooks.ibm.com/rss.html>

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 nesta publicação em outros países. Consulte um representante IBM local para obter informações sobre produtos e serviços disponíveis atualmente em sua área. Qualquer referência a produtos, programas ou serviços IBM não significa que apenas produtos, programas ou serviços IBM possam ser utilizados. Qualquer produto, programa ou serviço funcionalmente equivalente, que não infrinja nenhum direito de propriedade intelectual da IBM, poderá ser utilizado em substituição a este produto, programa ou serviço. Contudo, a avaliação e verificação da operação de qualquer produto, programa ou serviço não IBM são de responsabilidade do Cliente.

A IBM pode ter patentes ou solicitações de patentes pendentes relativas a assuntos tratados nesta publicação. O fornecimento desta publicação não concede ao Cliente direito algum sobre tais patentes. Pedidos de licença devem ser enviados, por escrito, para: *Gerência de Relações Comerciais e Industriais da IBM Brasil Av. Pasteur, 138-146 Botafogo Rio de Janeiro, RJ CEP 22290-240*

O parágrafo a seguir não se aplica a nenhum 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 NÃO SE LIMITANDO, ÀS 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 desse produto IBM e o uso 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.

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 iguais em sistemas geralmente disponíveis. Além disso, algumas medidas podem ter sido estimadas por extrapolação. Os resultados reais podem variar. Os usuários deste documento devem verificar os dados aplicáveis para seu ambiente específico.

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

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

LICENÇA DE COPYRIGHT:

Estas informações contêm programas de aplicativos de amostra na linguagem fonte, ilustrando as técnicas de programação em diversas plataformas operacionais. O Cliente pode copiar, modificar e distribuir estes programas de amostra sem a necessidade de pagar à IBM, com objetivos de desenvolvimento, utilização, marketing ou distribuição de programas aplicativos em conformidade com a interface de programação de aplicativo para a plataforma operacional para a qual os programas de amostra são criados. Esses exemplos não foram testados completamente em todas as condições. Portanto, a IBM não pode garantir ou implicar a confiabilidade, manutenção ou função destes programas.

Este documento, REDP-4913-00, foi criado ou atualizado em November 20, 2013.



Marcas Registradas

IBM, o logotipo IBM e ibm.com são marcas ou marcas registradas da International Business Machines Corporation nos Estados Unidos e/ou em outros países. Estes e outros termos com marca registrada IBM são marcados em sua primeira ocorrência nestas informações com o símbolo (® ou ™), que indica marcas registradas ou marcas registradas de direito consuetudinário nos Estados Unidos de propriedade da IBM no momento em que estas informações foram publicadas. Tais marcas registradas também podem ser marcas registradas ou de direito consuetudinário em outros países. Uma lista atual das marcas registradas IBM está disponível na Web em <http://www.ibm.com/legal/copytrade.shtml>



Os termos a seguir são marcas registradas da International Business Machines Corporation nos Estados Unidos e/ou em outros países:

IBM®

Redbooks®

Redguide™

Redbooks (logotipo) ®

SPSS®

Os termos a seguir são marcas registradas de outras empresas:

Outros nomes de empresas, produtos e serviços podem ser marcas registradas ou marcas de serviço de terceiros.