

Vitamina D

O Vitamina D é um produto para Tratamento e Limpeza de Dados voltado às áreas de CRM, Database Marketing e Data Warehouse. O Vitamina D foi desenvolvido para ser uma ferramenta global de fácil implementação e imediata produtividade.

Investimentos na ordem de um milhão de dólares em um valioso trabalho de desenvolvimento, ampliaram as funcionalidades e desempenho do produto. Seguindo rígidos padrões de Controle de Qualidade, o Vitamina D é hoje um sucesso comprovado com dezenas de Implementações nas mais diversas empresas.

A especialidade do Vitamina D é o Tratamento e Cruzamento de arquivos e cadastros genéricos. Possui rotinas para Normalização e Padronização, Identificação de Tipos de Pessoa, Correção e Separação dos Componentes de Nome e Endereço, Acentuação, Validação de DDDs, Emails e Prefixos Telefônicos, Validação de CPF, CNPJ e Inscrição Estadual, Tratamento de Caixa Alta/Baixa para processos de personalização, entre outras importantes rotinas de tratamento. Todo o ambiente de desenvolvimento das scripts de tratamento é totalmente integrado ao módulo VBScript, possibilitando aos usuários, infinitas possibilidades de parametrização e processamento das mais sofisticadas regras e tabelas de decisão.

O Vitamina D é disponibilizado nas versões:

Batch: Windows NT, 2000 e XP

Online: A versão API (Active X/COM) possui o acervo completo dos melhores componentes de qualidade de dados utilizados pelo Vitamina D. Esses componentes poderão ser utilizados pelos usuários dentro de suas próprias aplicações windows. A versão API é plug and play com processos de Intranet ou Extranet.

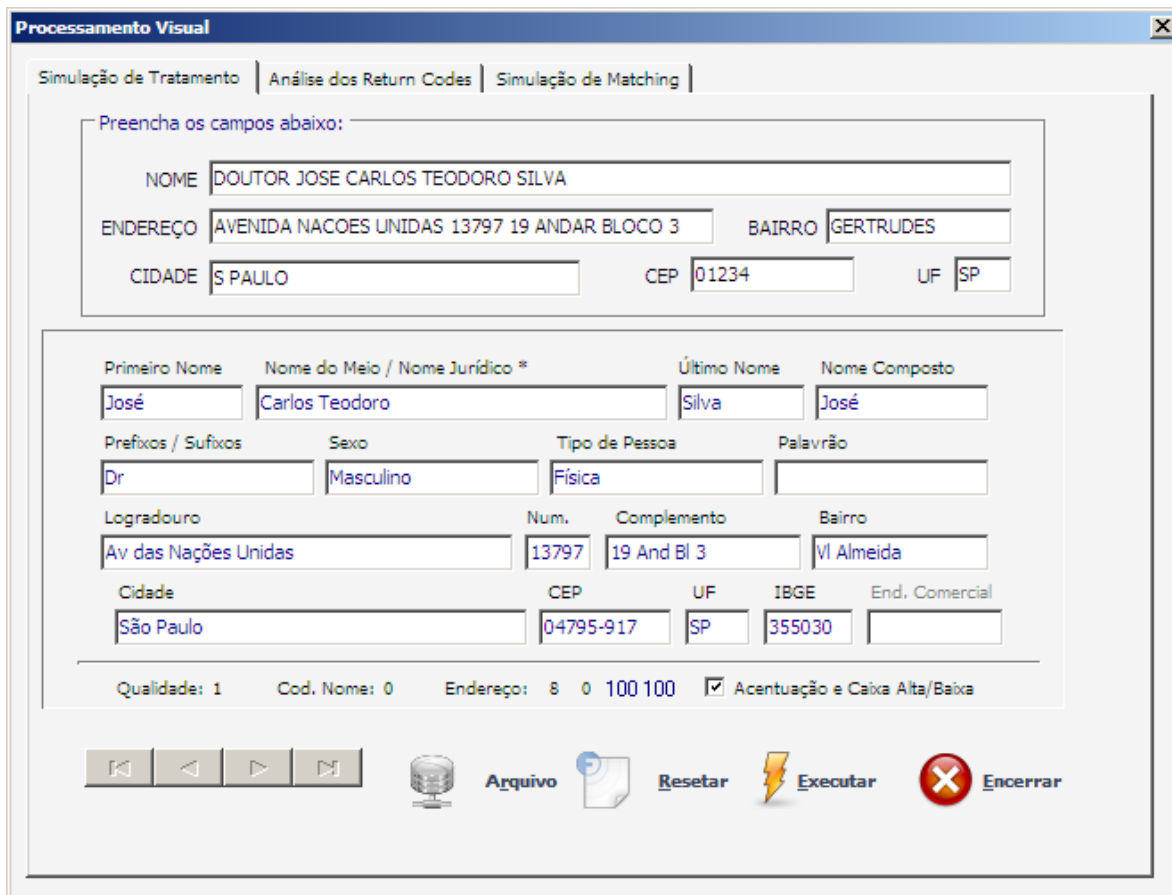
Recursos

O **Vitamina D** foi desenhado para ser uma ferramenta global de fácil implementação e imediata produtividade. Para melhor se adequar as necessidades de cada projeto, o **Vitamina D** foi dividido em módulos:

Módulo de Conversão: tem por finalidade Adaptar ou Converter um determinado arquivo para o padrão **Vitamina D**. Nesse módulo serão criados todos os campos de return-code que armazenarão os resultados e controles de um determinado tratamento.

O Vitamina D possui acesso nativo aos formatos SQL Server, Oracle, Sybase, Informix, DB2, ACCESS e DBF.

Tela 1 – Recursos do Tratamento Visual (on the fly):



The screenshot shows the 'Processamento Visual' window with the following data entered:

Simulação de Tratamento | Análise dos Return Codes | Simulação de Matching

Preencha os campos abaixo:

NOME: DOUTOR JOSE CARLOS TEODORO SILVA
 ENDEREÇO: AVENIDA NACOES UNIDAS 13797 19 ANDAR BLOCO 3 BAIRRO: GERTRUDES
 CIDADE: S PAULO CEP: 01234 UF: SP






Primeiro Nome	Nome do Meio / Nome Jurídico *	Último Nome	Nome Composto
José	Carlos Teodoro	Silva	José

Prefixos / Sufixos	Sexo	Tipo de Pessoa	Palavrão
Dr	Masculino	Física	

Logradouro	Num.	Complemento	Bairro
Av das Nações Unidas	13797	19 And Bl 3	Vl Almeida

Cidade	CEP	UF	IBGE	End. Comercial
São Paulo	04795-917	SP	355030	

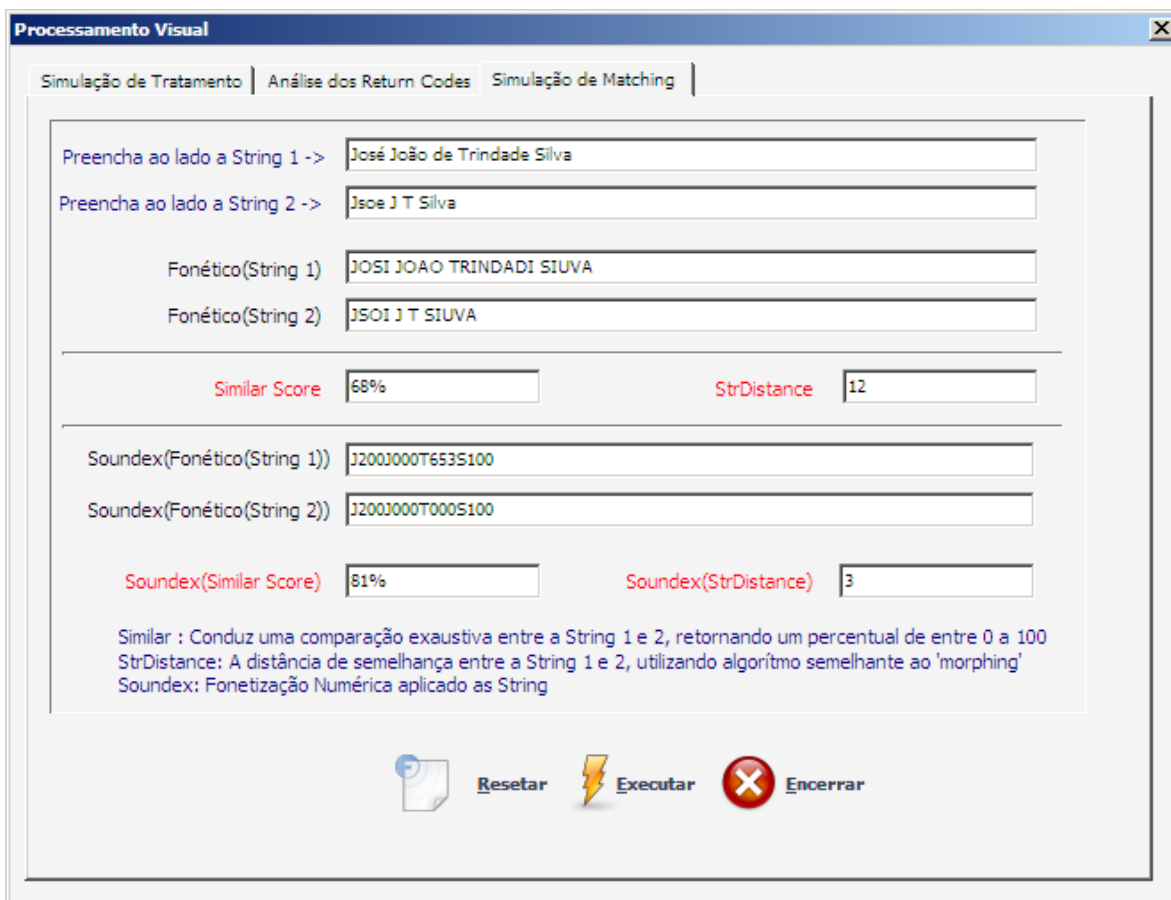
Qualidade: 1 Cod. Nome: 0 Endereço: 8 0 100 100 Acentuação e Caixa Alta/Baixa



 Arquivo
 
 Resetar
 
 Executar
 
 Encerrar

Módulo de Diagnóstico: Através do Módulo de Diagnóstico, pode-se analisar e diagnosticar o conteúdo e formato de um determinado arquivo. O diagnóstico é uma fase que obrigatoriamente precede o próprio tratamento. A visualização dos dados é realizada através de estatísticas de frequência e preenchimentos dos campos especificados pelo usuário. Exemplos:

- Descrição do Layout
- Tipos de campos e quantidade de registros
- Distribuição de preenchimento (análise dos missing)
- Frequência simples de um campo
- Frequência cruzada entre dois campos

Tela 2 – Recursos da Simulação de Matching (on the fly):



Processamento Visual

Simulação de Tratamento | Análise dos Return Codes | **Simulação de Matching**

Preencha ao lado a String 1 -> José João de Trindade Silva

Preencha ao lado a String 2 -> Jsoc J T Silva

Fonético(String 1) JOSI JOAO TRINDADI SIUVA

Fonético(String 2) JSOI J T SIUVA

Similar Score 68% StrDistance 12

Soundex(Fonético(String 1)) J200J000T653S100

Soundex(Fonético(String 2)) J200J000T000S100

Soundex(Similar Score) 81% Soundex(StrDistance) 3

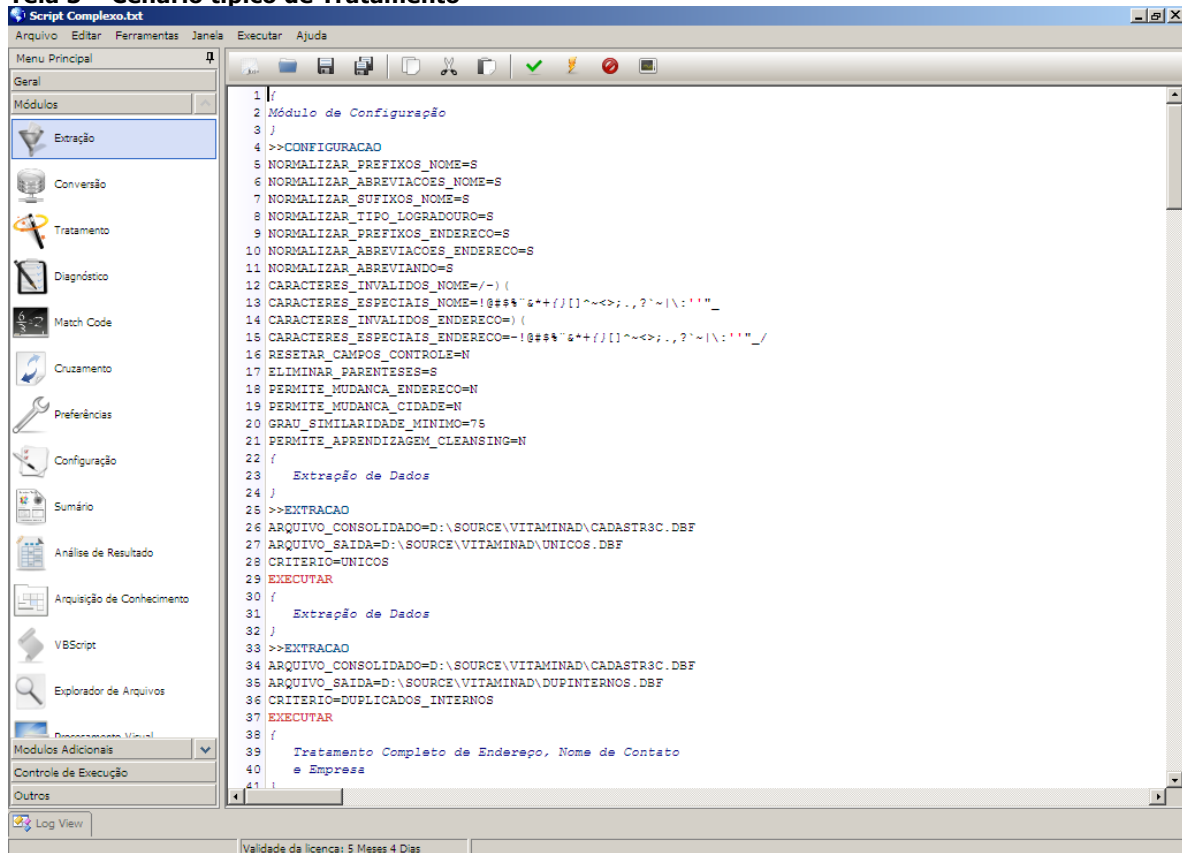
Similar : Conduz uma comparação exaustiva entre a String 1 e 2, retornando um percentual de entre 0 a 100
 StrDistance: A distância de semelhança entre a String 1 e 2, utilizando algoritmo semelhante ao 'morphing'
 Soundex: Fonetização Numérica aplicado as String

Resetar Executar Encerrar

Módulo de Tratamento: O principal módulo do **Vitamina D**. Aqui é realizado o Tratamento de Dados propriamente dito. Entre as facilidades principais:

- Identificação de tipo de Pessoa Física e Jurídica
- Determinação de sexo para Personalização e Nomes Compostos
- Normalização de Empresa, Nome (Funcionário) e Cargos
- Normalização / Separação dos componentes de endereço
- Correção de endereço e CEP
- Identificação de endereços comerciais e aplicação do código do IBGE
- Identificação de palavras, Inclusão e exclusão de acentos, e caixa alta/baixa
- Validação de Email/Telefones
- Validação de CPF/CNPJ/Inscrição Estadual
- Relatórios Sumarizados ou Detalhados

Tela 3 – Cenário típico de Tratamento



```

1 |
2 Módulo de Configuração
3 |
4 >>CONFIGURACAO
5 NORMALIZAR_PREFIXOS_NOME=S
6 NORMALIZAR_ABBREVIACOES_NOME=S
7 NORMALIZAR_SUFIXOS_NOME=S
8 NORMALIZAR_TIPO_LOGRADOURO=S
9 NORMALIZAR_PREFIXOS_ENDERECO=S
10 NORMALIZAR_ABBREVIACOES_ENDERECO=S
11 NORMALIZAR_ABBREVIANDO=S
12 CARACTERES_INVALIDOS_NOME=/-/)
13 CARACTERES_ESPECIAIS_NOME=!@##%*'+/{}[]^~<>.;:.,?'~\:\:'''_
14 CARACTERES_INVALIDOS_ENDERECO=(
15 CARACTERES_ESPECIAIS_ENDERECO=-!@##%*'+/{}[]^~<>.;:.,?'~\:\:'''_/_
16 RESETAR_CAMPOS_CONTROLE=N
17 ELIMINAR_PARENTESES=S
18 PERMITE_MUDANCA_ENDERECO=N
19 PERMITE_MUDANCA_CIDADE=N
20 GRAU_SIMILARIDADE_MINIMO=75
21 PERMITE_APRENDIZAGEM_CLEANING=N
22 {
23     Extrapão de Dados
24 }
25 >>EXTRACAO
26 ARQUIVO_CONSOLIDADO=D:\SOURCE\VITAMINAD\CADASTR3C.DBF
27 ARQUIVO_SAIDA=D:\SOURCE\VITAMINAD\UNICOS.DBF
28 CRITERIO=UNICOS
29 EXECUTAR
30 {
31     Extrapão de Dados
32 }
33 >>EXTRACAO
34 ARQUIVO_CONSOLIDADO=D:\SOURCE\VITAMINAD\CADASTR3C.DBF
35 ARQUIVO_SAIDA=D:\SOURCE\VITAMINAD\DUPLICADOS.DBF
36 CRITERIO=DUPLICADOS_INTERNOS
37 EXECUTAR
38 {
39     Tratamento Completo de Endereço, Nome de Contato
40     e Empresa
41 }
  
```

Tela 4 – Parametrização do processamento:

Algumas características do Kernel de Cleansing na versão atual

- O algoritmo de gerenciamento e administração de conteúdo foi melhorado. Hoje, o Vitamina D processa tokens e realiza análises sintáticas de uma forma muito mais otimizada.
- Melhor avaliação no Parser de Endereço. (redução de matches falsos)
- O acesso à knowledge base foi otimizado e agora consegue processar e interpretar os dados mais eficientemente. Este ganho de velocidade está sendo utilizado para adicionar mais conhecimento durante o processo de limpeza.
- Melhor otimização na procura de elementos das tabelas.

Módulo de Match-Code: Através do módulo de Match Code são geradas todas as TAGs para utilização durante toda a etapa de cruzamento. A partir das parametrizações informadas, o **Vitamina D** pode se adaptar a virtualmente qualquer cenário e perfil de cadastro ou base de dados. Uma função particularmente útil é a priorização de campos, registros e listas através de pesos. Por exemplo: Um cadastro onde uma determinada empresa apresenta inúmeras duplicidades. Como selecionar qual executivo deverá ser contatado? Uma opção é parametrizar como qualificador o campo cargo. Através dessa função, o **Vitamina D** pode selecionar somente aqueles executivos de mais alta cargo hierárquico, quando ocorrer duplicação por nome de empresa.

Match Code

Selecione o Arquivo: C:\Arquivos de programas\VitaminaD4\BRA\Cadastr2c.dbf





Match Code: P. Física 0 Kb 29/7/2010 12:45:31 Arquivo Convertido

Genérico
 House Hold
 P. Física
 P. Jurídica
 Total

	Tipos	Integral	Sub TAG	Definição das Partículas		
				Inicial	Central	Final
Empresa:	<input type="checkbox"/>	<input type="checkbox"/>	7	4	0	3
Nome: D_N1	<input type="checkbox"/>	<input type="checkbox"/>	8	4	1	3
Logradouro/Num: D_G2	<input type="checkbox"/>	<input type="checkbox"/>	11 *	1	1	4
Complemento:	<input type="checkbox"/>	<input type="checkbox"/>	8	8		
CEP: D_G6	<input type="checkbox"/>	<input type="checkbox"/>	3	3		
Campo Extra 1:	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0	0
Campo Extra 2:	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0	0

TAG: D_TAG Tamanho -> 22 * - Inclui 5 bytes do Número do Logradouro

Método TAG: D_MET Composição -> 000413.11403.000.000

 Resetar
  Script
  Encerrar
  Ajuda
 Arquivo Previamente Tratado

Módulo de Cruzamento: Nessa fase é realizado o Merge & Purge. O **Vitamina D** através das Tags criadas no módulo de Match-code, utilizando os algoritmos de deduplicação, realiza as correlações e identificações de registros duplicados. Os resultados do Merge & Purge incluem:

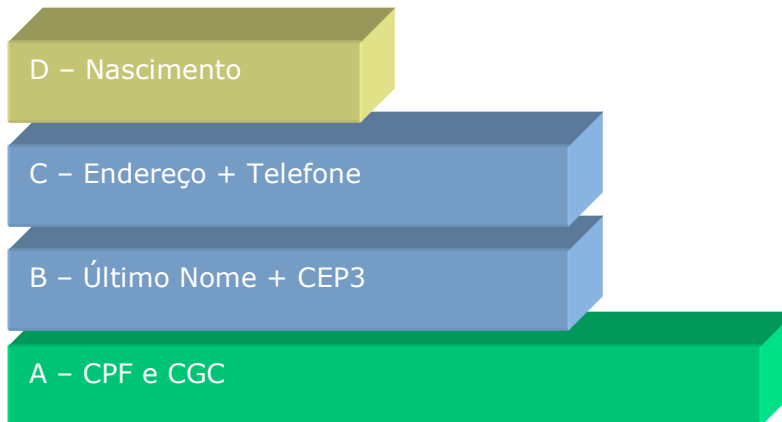
- Marcação da base
- Relatórios de Input / Output
- Rejeitados
- Multi-buyers Internos e Externos
- Crosstab entre fontes (listas)
- Categorias livres

Funcionalidade das Match Keys Ponderadas:

Matches Keys ponderadas, utilizam diversas variáveis para determinar através de regras (matching rules), quando um registro deve ser considerado duplicado ou único.

Através de pesos (Ponderações) serão definidas a força de uma determinada condição na Tabela de Decisão.

Exemplo: Função $f(Ax7, Bx3, Cx3, Dx2)$



Módulo de Extração: A etapa final do Tratamento de Dados e do Cruzamento. Através das facilidades incluídas nesse módulo, pode-se extrair virtualmente qualquer registro de um cadastro, dado um critério pré definido de uma variável ou resultado do processamento. Critérios disponíveis:

- Únicos (Registros que não apresentaram qualquer duplicidade)
- Duplicados Internos
- Duplicados Externos
- Registros Multibyers
- Deduplicados
- Sem critério de Segmentação
- Por Código de Lista

Performance

Vitamina D

O que torna o **Vitamina D** ainda mais poderoso, é a forma como suas funções e comandos são acessados. Através da Linguagem Script, pode-se definir Interativamente ou diretamente através do Editor, todos os comandos e etapas necessários de um tratamento. Uma vez o cenário do tratamento definido, pode-se executá-lo instantaneamente, ou ainda em batch.

Utilizando os recursos da Script de Tratamento, podem ser encadeados ilimitados diagnósticos e tratamentos, que serão realizados serialmente. Um dos recursos mais avançados e que trazem a melhor eficácia ao **Vitamina D**, é o processo de Execução de Scripts em batch.

Como ponto forte na automação e produtividade, a possibilidade de encadear processos de tratamento, faz o **Vitamina D** trocar informações com outras aplicações pré existentes do usuário.

O **Vitamina D Batch** pode ser executado no modo Runtime. Através da execução runtime, o **Vitamina D** pode ser utilizado em processos conjuntos com uma aplicação externa do usuário, sem a necessidade de instalação de nenhum outro componente (dll, ocx, api), além aqueles utilizados na aplicação do usuário.

Utilitários de Apoio

Auto Aprendizagem

O Knowledge Base do Vitamina D foi dividido em 2 categorias.

- A primeira delas, chamada de Base de Conhecimento Nativo, contém as regras fundamentais de tratamento e está disponível somente para leitura.
- A segunda categoria, chamada de TABLEBASE, é editável. Com a implementação desta função, o usuário pode ensinar ao aplicativo quando e o que deve ser feito diante dos mais diversos problemas de cleansing que aparecem no dia a dia.

Gerador de Query

O Gerador de Query na realidade, é um poderoso gerador de consultas e gerenciador de arquivos integrado. A função dele é justamente ser uma ferramenta para que o usuário possa conhecer melhor o conteúdo de um determinado arquivo genérico, utilizando todas as facilidades do SQL.

Construtor de Comandos

Através deste utilitário, o usuário encontrará múltiplas rotinas para controlar seus problemas ocasionais. Todos estes comandos são integrados com o IDE do Vitamina D (permitindo entre outros recursos a execução de Processos em Batch, Multi-Tarefa e Acesso Nativo a bases SQL). Entre os Comandos Principais, foram incluídos a Correção CPF/CGC, assim como a Manutenção de Telefone (Código da área e Correção do Prefixo). Mais de 50 comandos de finalidades gerais dentro do Command Construction Utility.

CEP Eletrônico

O CEP Eletrônico é um guia de Ruas e Ceps para todo o território nacional. Porém recursos técnicos altamente específicos o diferenciam dos demais produtos do mercado. O software permite a procura por sentenças ou Ceps em qualquer campo do Cadastro do Correio, em qualquer posição e em qualquer ordem, além de contar com operadores que facilitam a orientação de uma determinada pesquisa.

Frequently Asked Questions

Vitamina D - BR

1. Qual a finalidade do Vitamina D ? E porque o nome 'Vitamina D' ?

*O **Vitamina D**® é um produto nacional, desenhado para processos de Higienização de Dados, Database Marketing e Data Warehouse.*

Com o Vitamina D, você pode vitaminar e melhorar a saúde de todos os cadastros utilizados em operações de Marketing Direto, assim como, é possível enriquecer os processos de preparação e suporte de bases de clientes e prospects.

Em geral, todo o ciclo de Tratamento de Dados realizado pelo Vitamina D, pode ser dividido em quatro grandes modalidades: Diagnóstico, Higienização, Deduplicação e Estatísticas

O compromisso da Harte-Hanks com o lançamento da nova versão do Vitamina D é simples: satisfazer cada vez mais e melhor as necessidades dos seus consumidores, aumentar a retenção, fidelização, e assim seu lifetime value; Reduzir custos de produção e postagem através da eliminação dos registros duplicados e inválidos de seus arquivos; Maximizar os descontos por volume de malas-diretas; Incrementar a qualidade dos seus dados de nome e endereço, permitindo maior acuracidade na segmentação de seus prospects.

2. Me dê alguns exemplos de utilização do Vitamina D...

O Vitamina D foi desenhado para ser Plug and Play com os processos de Tratamento de Dados de sua empresa. O Vitamina D é oferecido ao usuário como um pacote preparado para Higienizar Dados imediatamente, já em sua primeira execução.

O Vitamina D pode ser utilizado em operações de Marketing Direto, maximizando recursos, eliminando gastos desnecessários com informações incorretas, como, registros com Endereços Inválidos ou duplicados.

Em operações com produtos financeiros, o Vitamina D pode detectar e remover ofertas múltiplas para o mesmo indivíduo. O Vitamina D pode ser utilizado de uma forma muito eficaz, durante a criação e manutenção de Cadastros Únicos (Relationship Matching)

No Data Warehouse pode eficientemente criar links de informações correlatas em ambientes transacionais heterogêneos, além evidentemente, padronizar, depurar e limpar qualquer datamart.

Os Executivos de Planejamento Estratégico, seja em Data Mining ou Análises Estatísticas podem utilizar o Vitamina D, para identificar e remover elementos duplicados que frequentemente levão a conclusão incorreta de um determinado estudo de mercado.

O Vitamina D é muito fácil de usar. Graças aos recursos de Script de Tratamento, possui uma incrível flexibilidade para utilização e tratamento de diversos tipos de layout.

3. Minha empresa tem uma operação de Marketing Direto muito sólida. Que benefícios teria em utilizar o Vitamina D, por exemplo, em minhas campanhas de Mala Direta ?

O Vitamina D é um produto que paga-se sozinho. Ou seja, dependendo do porte de sua operação, o ROI é atingido em uma única campanha.

O Vitamina D possui as melhores facilidades do mercado, para tratar uma operação de Mala Direta, nos mais incríveis detalhes. Dentre as principais características, pode-se destacar:

- *Identificação de Pessoa Física e Jurídica (Você não corre o risco de personalizar uma mala direta, chamando uma empresa de Prezada Senhora !)*
- *Determinação de Sexo para Personalização (Além de nunca mais precisar utilizar o famoso artifício Senhor(a), você pode personalizar o corpo do texto de sua peça publicitária, conforme o sexo de seu leitor)*
- *Nomes Compostos (Chame o Sr. José, pelo seu nome composto)*
- *Normalização de Empresa e Nome*
- *Normalização / Separação dos componentes de Endereço*
- *Correção de Endereço e CEP (Não jogue fora seu dinheiro. Tenha o endereço de seu público, completamente revisado e corrigido)*
- *Inclusão de Acentos e tratamento de letras (Caixa Alta/Baixa)*
- *Tratamento do Telefone e CPF/CNPJ e Emails*
- *Plug and Play com processos de tratamento já existentes*
- *Schedule de Jobs (via Scripts)*
- *Performance excepcional*

O Vitamina D é a melhor ferramenta para garantir um excelente retorno em sua Mala Direta.

4. Qual a plataforma recomendada para sua utilização ?

O Vitamina D foi desenhado para aproveitar ao máximo os recursos e potencialidades do ambiente Windows. O Vitamina D roda em Windows 95, 98, NT, 2000 e XP. Não existe uma configuração recomendada para o Tratamento de Dados com o Vitamina D. Porém quanto mais rápido for o processador, a controladora de acesso ao disco, assim como, a quantidade de memória RAM disponível, mais rápido e eficiente será o Tratamento de Dados.

5. Tenho cerca de 2 milhões de registros para tratar. O Vitamina D daria conta do recado ? Quais são os limites do Vitamina D ?

Sem problemas. O Vitamina D seria completamente capaz de tratar 2 milhões de registros. Dê uma olhada nos limites máximos do Vitamina D:

- *Registros por tabela: 1 bilhão*
- *Bytes por tabela: 2 Gigabytes*
- *Caracteres por Registro: 65.000*
- *Campos por Registro: 255*

Nos Laboratórios de Testes do Vitamina D, durante um tratamento de dados COMPLETO, utilizando-se um Pentium IV 1Giga, Windows 2000, 256 MbRAM, U/W SCSI-3, atingiu-se a marca média de 20 registros tratados por segundo !!!

6. Quais as características do Módulo de Merge & Purge do Vitamina D ?

Dentre as principais características da etapa de Cruzamento e Deduplicação do Vitamina D, pode-se destacar:

- *Criação completamente parametrizada das Match Keys*
- *Cruzamento dos Dados (com layouts únicos ou diferentes entre si)*
- *Identificação de Duplicados e Multibyters*
- *Deduplicação por até 256 fontes diferentes*
- *Extração de Não Duplicados/Duplicados (por fonte x Multibyters)*
- *Deduplicação por Matchs Ponderadas*
- *Relatórios Analíticos (com Samples)*
- *Crosstab de Duplicidade por Fonte*
- *Estatísticas Completas*

7. O Vitamina D Batch é um daqueles softwares onde são necessárias semanas de consultoria de implantação para começar a 'funcionar' ?

Não. Ao contrário de seus principais concorrentes no Brasil, o Vitamina D é um software de utilização imediata. Logo após a instalação do Vitamina D em seu equipamento (processo esse que demora não mais que 10 minutos), o Vitamina D já estará pronto para realizar um Tratamento de Dados.

Lembre-se, O Vitamina D foi desenhado para ser Plug and Play com os processos de Tratamento de Dados de sua empresa. O Vitamina D é oferecido ao usuário como um pacote preparado para Higienizar Dados, imediatamente em sua primeira execução. Não será necessário a criação de programas ou horas de consultoria de 'adaptação'. Tudo o que você terá que fazer, é dizer ao Vitamina D, o que ELE terá que fazer...

Além de ser intuitivo, com interface totalmente Gráfica, o Vitamina D possui um excelente manual do usuário, com mais de 200 páginas, que são extratadamente didáticas, ilustrando todas as características do produto. O próprio Help Online, acompanha 7 lições em forma de Tutoriais, para quem está interessado em conhecer o Produto.

8. Em que linguagem foi desenvolvido o Vitamina D ?

O Vitamina D foi desenvolvido 100% em Delphi (RAD Studio 2010).

9. Quem é a Harte Hanks ?

A Harte-Hanks Data Technologies é uma empresa americana, sediada em Boston, MA líder mundial em serviços de Database Marketing e Datacleansing. Com mais de 5.000 profissionais e 30 anos de experiência na área, a empresa sempre manteve um crescimento contínuo firmando uma presença inovadora nos EUA, Canadá, Europa, Ásia e América Latina. Seus produtos têm como principal diferencial a portabilidade, escalabilidade, alta performance, confiabilidade e facilidade para análises de negócios.

Copyright © 2010 Harte Hanks Data Technologies. All rights reserved.