



Escapando da armadilha Datamart:

CONSOLIDANDO DATAMARTS

Por W H Inmon

Olhe ao redor da organização de TI moderna, o que você encontra? Você encontra datamarts em todos os lugares. Em alguns casos pode-se encontrar centenas deles. E quantos estão realmente sendo utilizados? Na maioria dos casos, apenas um punhado está sendo utilizado.

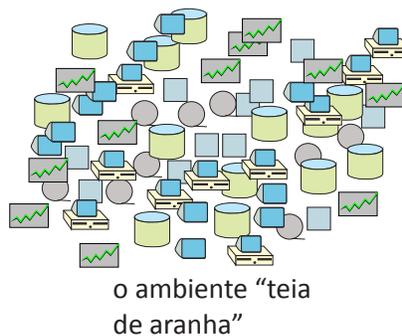
Este dilema é dispendioso e desconcertante. Certamente uma organização pode fazer melhor que isso.

Este white paper examinará como e por que as organizações chegaram a estas ocorrências e o que pode ser feito a esse respeito. Compreender a história de como uma organização chegou a esse dilema é o primeiro passo para descobrir como resolver os problemas de excesso de datamarts.

UMA BREVE HISTÓRIA

Quase todas as organizações começaram a vida com uma série de aplicativos operacionais chamados de “ambiente operacional”. Alguns desses aplicativos operacionais foram construídos internamente e mantidos ao longo dos anos. Outros foram adquiridos a partir de fornecedores de terceiros. E outros, ainda, foram herdados de aquisições corporativas e fusões. No entanto, foram adquiridos, e estes aplicativos operacionais executam os negócios diários da organização.

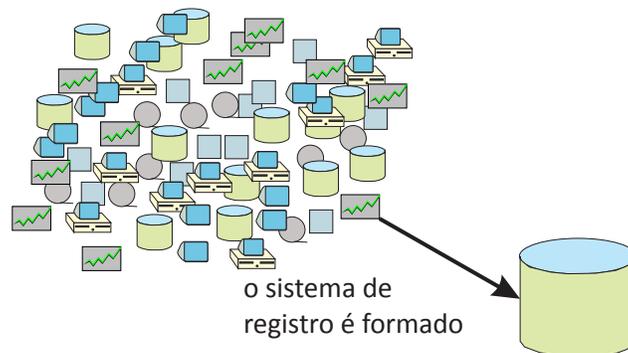
A figura 1 ilustra estes aplicativos operacionais.



Os aplicativos operacionais foram o ponto de partida para quase todas as organizações. A partir deste ponto as organizações tomaram um dos dois caminhos. O primeiro caminho a ser discutido é aquele que pode ser chamado de “única versão da verdade”. No caminho da “única versão da verdade”, foi reconhecido que entre os muitos aplicativos operacionais não havia um

entendimento definitivo dos dados corporativos básicos. Uma empresa de seguros não sabia quantas apólices possuía. Um varejista de pedidos por correio não sabia quem eram seus clientes. Um banco não era capaz de conectar as diferentes contas de um mesmo cliente. Uma empresa de seguros não entendia a magnitude do risco que enfrentava, e assim por diante.

As empresas podiam observar e manipular dados operacionais a níveis detalhados até o cerne de seu conteúdo, mas não eram capazes de examinar a empresa em sua totalidade e extrair qualquer informação. Como um recurso para essa proposição, foi criado um data warehouse com base no estilo de arquitetura da única versão da verdade. O chamado “sistema de registro” (ou “única versão da verdade”) proveio dos melhores dados de origem que a empresa possuía. A figura 2 ilustra o estilo de data warehouse da única versão da verdade.



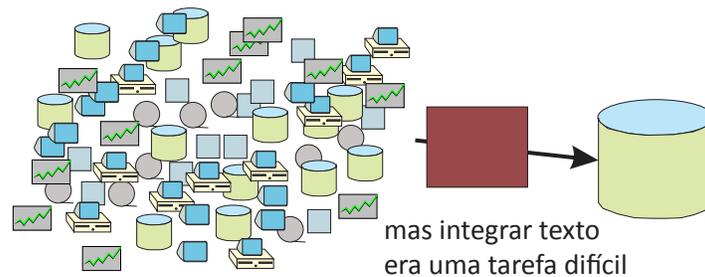
INTEGRANDO DADOS OPERACIONAIS

No estilo de construção de um data warehouse da única versão da verdade, as empresas de armazenamento de dados descobriram que, como estavam integrando dados de muitas fontes diferentes, a integração de dados operacionais era uma tarefa difícil. Existiam muitas razões pelas quais integrar uma grande quantidade de dados operacionais distintos era difícil:

- Os aplicativos operacionais frequentemente não eram documentados devidamente.
- Os aplicativos operacionais foram construídos em tecnologias desatualizadas.
- Os aplicativos operacionais foram construídos a partir de sistemas de múltiplas fontes, onde tais sistemas nunca foram projetados para trabalhar em conjunto.

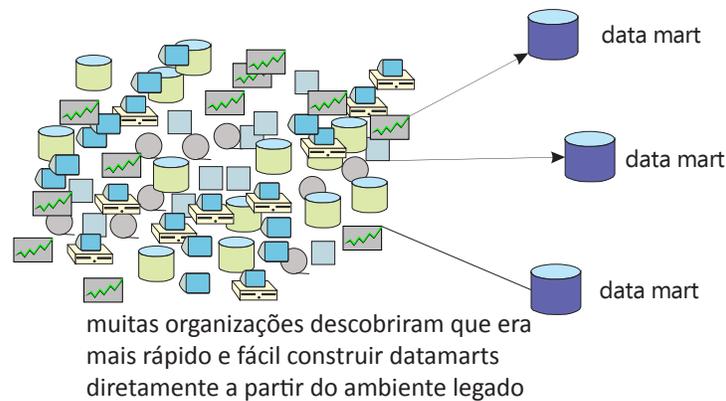
Na verdade, existiram várias razões pelas quais as organizações simplesmente não quiseram enfrentar a desordem gerada pela combinação de inúmeros e distintos sistemas operacionais.

A figura 3 mostra o problema causado pela integração do sistema operacional.



A integração de dados de sistemas operacionais era um esforço tal, que muitas empresas simplesmente não tentariam a tarefa. Assim surgiu o segundo caminho para muitos datamarts que existem atualmente. O segundo caminho pode ser chamado de abordagem “dimensional” para o armazenamento de dados. Nessa abordagem, as organizações não construíam um data warehouse de modo algum. Ao invés disso, as organizações construíam uma coleção de datamarts que era provida diretamente por aplicativos operacionais. Muitos fornecedores de software favoreceram a abordagem “dimensional” porque com ela o fornecedor de software poderia vender seu produto imediatamente, enquanto que na abordagem da “única versão da verdade”, o fornecedor de software precisava esperar até que fosse construído um data warehouse. Uma vez que os fornecedores de software queriam fazer todo o possível para encurtar o ciclo de vendas, a abordagem “dimensional” foi amplamente favorecida.

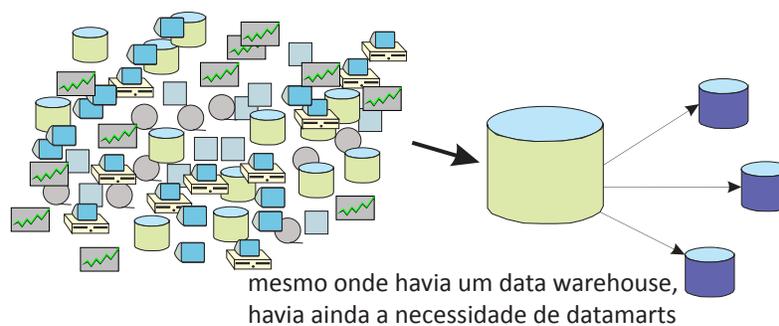
A figura 4 ilustra a abordagem “dimensional” que consistia de aplicativos operacionais e uma coleção de datamarts construídos diretamente a partir do ambiente de sistemas operacionais.



Mas, mesmo com a abordagem “única versão da verdade” (Single Version of the Truth ou SVOT) para armazenamento de dados, havia a necessidade de datamarts. Com essa abordagem, os datamarts eram providos diretamente a partir do data warehouse, onde a única versão da verdade residia.

Assim, se a organização escolhesse a abordagem “única versão da verdade” ou a abordagem “dimensional”, o resultado era uma coleção de datamarts.

A figura 5 ilustra a abordagem “única versão da verdade” para armazenamento de dados e os datamarts associados que fluem do data warehouse.

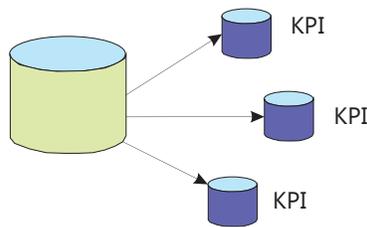


KPIs

De qualquer forma, a base conceitual do datamart foi um KPI (Principal Indicador de Desempenho). Cada organização possui o seu KPI. Os KPIs da organização indicam determinadas medidas de como os negócios de uma empresa estão se saindo. Algumas organizações tinham índice de pedidos para faturar. Outras organizações tinham a medida de captura e proteção de participação no mercado. Outras empresas decidiram-se pelo aumento das receitas. E outras, ainda, decidiram-se pela rentabilidade. Cada empresa tinha seu próprio conjunto de

medidas que refletiam seu funcionamento e direção. E, geralmente, estes KPIs tinham como base os dados encontrados nos datamarts.

A figura 6 mostra que os datamarts suportam os KPIs da empresa.

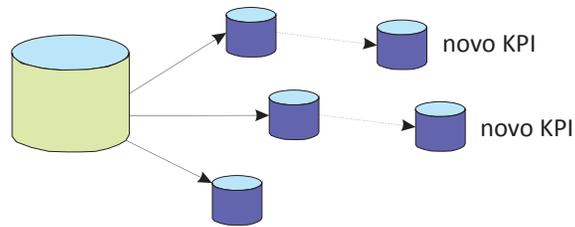


O gerenciamento ficou satisfeito ao descobrir que os KPIs poderiam ser suportados pelos datamarts. E independentemente de a organização ter escolhido uma abordagem “única versão da verdade” para armazenamento de dados e datamarts, ou uma abordagem de modelo “dimensional” para datamarts, existiam datamarts que serviam de base para os KPIs. Logo, KPIs de suporte a finanças, vendas, marketing, contabilidade, gerenciamento, engenharia e outras funções começaram a surgir.

Mas então aconteceu algo peculiar. O gerenciamento não estava satisfeito com o datamart logo após este ter sido construído. Havia muitas razões pelas quais o gerenciamento não estava satisfeito com seus datamarts, mas talvez a razão mais importante fosse que os KPIs continuavam mudando. À medida que os negócios mudavam, surgia a necessidade de um novo conjunto de KPIs. Em determinado momento, uma organização estava em modo de expansão, onde a renda e a participação no mercado eram tudo. Um conjunto de KPIs refletia a necessidade de aumentar a renda e o número de clientes. No dia seguinte, a empresa acordava para descobrir que estava em modo de rentabilidade. Os KPIs relativos a despesas e rentabilidade tornaram-se devastadores. Em um outro dia, a empresa acordava para descobrir que uma nova legislação havia sido aprovada e que um novo conjunto de dados era necessário. E no próximo, a empresa acordava com a nova e inesperada concorrência.

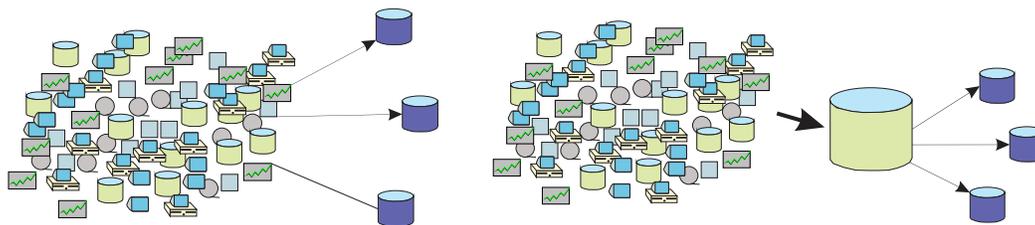
Com o passar do tempo os negócios mudaram e, por conta disso, havia uma necessidade constante de evolução para novos conjuntos de KPIs.

A figura 7 mostra a necessidade constante de evolução para novos KPIs.



No caso do modelo “dimensional”, houve um fator adicional para a necessidade de novos KPIs. No modelo de abordagem “dimensional” para data warehouses e datamarts, as organizações esforçaram-se para encontrar quais dados eram os dados “certos”. Em determinado momento, um KPI se mostrava interessante, e no minuto seguinte outro KPI era visto dessa maneira. Mas, mesmo com essa abordagem, havia uma mudança constante nos datamarts e KPIs.

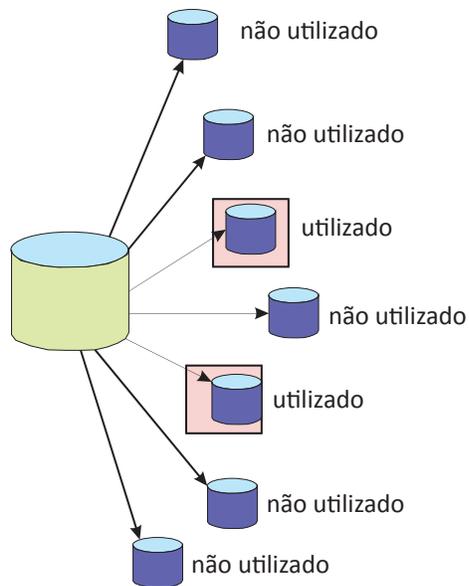
A figura 8 mostra que, tanto com a abordagem “única versão da verdade” como com o modelo “dimensional”, as condições de negócios estão sempre evoluindo, o que resultou na evolução dos datamarts e KPIs.



em ambas as abordagens existem datamarts

Independentemente do modelo escolhido pela organização, ela se conscientizou de que havia um número sempre crescente de datamarts. E, devido a esses datamarts terem aparecido repentinamente, apenas alguns estavam sendo realmente utilizados, enquanto muitos outros foram caindo em desuso.

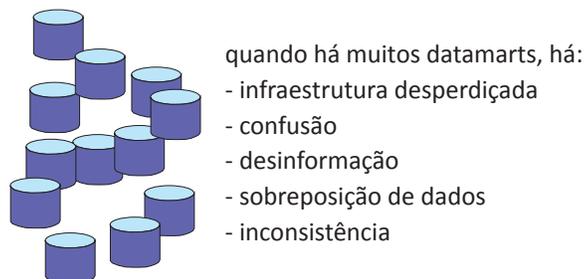
A figura 9 ilustra os datamarts utilizados e não utilizados que evoluíram.



Quando há grande quantidade de datamarts, surgem diversos problemas. Entre esses problemas estão:

- Pagamento dos recursos necessários para hospedar os datamarts e atualizá-los.
- Confusão sobre a qual datamart recorrer quando surge um problema.
- Confusão sobre como resolver as diferenças em valores quando dois ou mais datamarts fornecem informações contraditórias.
- Falta de consistência de um datamart em relação ao próximo.
- Entre outros.

Em resumo, quando há uma grande quantidade de datamarts, há potencial para muita confusão. E existem muitos recursos baseados em computador desperdiçados.



A figura 10 mostra o que acontece quando há uma grande quantidade de datamarts.

SOLUCIONANDO O DILEMA DOS DATAMARTS

TIPOS DIFERENTES DE DATAMARTS

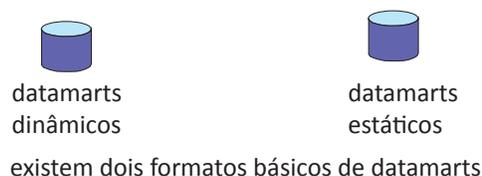
Existem muitas maneiras de classificar os datamarts. Uma delas é se eles foram criados dinamicamente ou estaticamente. Um datamart criado dinamicamente é aquele criado diretamente a partir dos dados de origem e a partir de dados que mudam com rapidez. Por exemplo, um banco pode criar um datamart contendo os valores de todas as contas bancárias a partir das 10h32 da manhã de terça-feira. Se o banco fosse recriar o datamart às 16h13, ele esperaria ter um conjunto diferente de valores em vigor.

Datamarts dinâmicos são frequentemente criados para uma análise transitória. Muitas vezes, esse datamart é destruído logo após ter sido criado. O tempo de vida total de um datamart dinâmico pode ser inferior a 10 ou 15 minutos.

Um datamart estático é bem diferente. Este datamart é construído para ter longa duração. Normalmente, ele é atualizado periodicamente. Por exemplo, a empresa pode medir o aumento ou a diminuição da participação no mercado mensalmente. Para isso, o datamart captura as informações mais atuais, mas também deixa uma trilha das medições em seu decorrer. Ao utilizar os dados históricos encontrados no datamart, a empresa pode controlar as alterações incrementais ao longo do tempo.

Quando uma organização finalmente descobre que há datamarts demais sendo suportados por ela, quase sempre está se referindo a datamarts estáticos e não dinâmicos.

A figura 11 ilustra datamarts estáticos e dinâmicos.

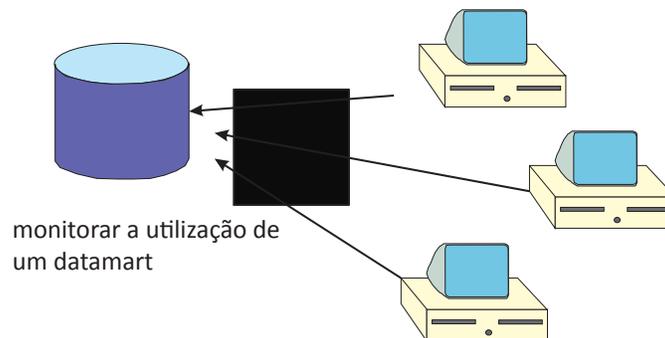


DETERMINANDO O USO

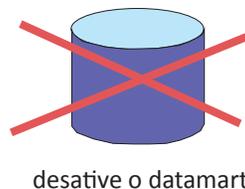
O primeiro passo para resolver a questão da utilização do datamart é determinar quais estão ou não sendo utilizados. A maneira mais simples de determinar isso é perguntar para as pessoas quais datamarts estão sendo utilizados. Embora questionar as pessoas sobre o que está acontecendo seja uma estratégia bastante educada, muitas vezes é uma medida imprecisa.

O modo correto para determinar a utilização de um datamart é monitorar sua atividade. O monitoramento do processamento de atividades de um datamart produz uma declaração bastante exata de sua utilização.

A figura 12 mostra que a utilização de um datamart pode ser mensurada.

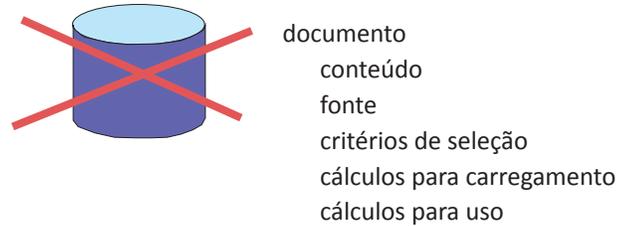


Ao monitorar a utilização de datamarts, o analista pode determinar precisamente quais estão ou não sendo utilizados. O segundo passo é parar de atualizar e de oferecer suporte aos datamarts que não estão sendo utilizados.



O próximo passo é, então, documentar o conteúdo e o processamento desses datamarts que não estão sendo utilizados. A intenção é descobrir qual KPI estava sendo atendido quando o datamart foi construído. Em alguns casos, a intenção do KPI é clara, em outros casos, não. De qualquer forma, a documentação da intenção do KPI é algo útil e criterioso de se fazer.

A figura 13 mostra a documentação dos datamarts não utilizados.



Após a identificação e documentação dos datamarts que não estão sendo amplamente utilizados, o próximo passo é rever a decisão de encerrar as atividades desses datamarts.

Existem muitas maneiras de encerrar um datamart. Uma delas é remover toda e qualquer infraestrutura associada ao datamart. Uma abordagem menos radical é simplesmente deixar a infraestrutura em vigor e desativar a etapa final de criação de um datamart.

Recomenda-se que a abordagem menos radical possível seja utilizada.

Depois que o datamart tiver sido desativado, o próximo passo é ficar atento e ouvir quais usuários finais irão reclamar. Se houver algum problema, o usuário final é direcionado a outro datamart, que esteja sendo ativamente suportado. O usuário final é incentivado a utilizar um dos datamarts mais ativamente utilizados e suportados.

A figura 15 mostra que o usuário final é incentivado a utilizar um datamart existente.



PAGANDO PELA INFRAESTRUTURA

Se o usuário final persistir em afirmar que o datamart desativado é de vital importância, então, ele é convidado a pagar por sua infraestrutura. Há um custo pelo período dos aspectos permanentes da infraestrutura e encargos contínuos para a reposição periódica de dados no datamart. Se o usuário final estiver disposto a pagar pela infraestrutura permanente e pela atualização do datamart, ele é, então, reintegrado. É nesse ponto que o conhecimento de uma abordagem não radical para a desativação do datamart fica evidente.

Se, após um período razoável, nenhum usuário final exigir que um datamart desativado seja reintegrado, toda a infraestrutura deste datamart não utilizado é removida.

O processo de monitoramento descrito é aquele que deve ser feito repetidamente e periodicamente. Não é um evento único.

Para os datamarts que estão sendo utilizados, é útil documentá-los.

Documentação de aspectos, tais como:

- A origem de dados prestes a fazerem parte do datamart.
- O algoritmo de seleção utilizado para dados prestes a fazerem parte do datamart.
- O algoritmo de cálculo dos dados que “fluem” para o datamart.
- O KPI (ou KPIs) que está(ão) na base do datamart.
- O período pelo qual o datamart deve ser mantido.
- O usuário-alvo do datamart.
- O tamanho projetado do datamart.
- As restrições de segurança associadas ao datamart.

O resultado final de responsabilizar-se por um programa de remoção de datamarts não utilizados sugerido, é para:

- Aperfeiçoar o processamento.
- Tornar o processamento e a infraestrutura mais acessíveis.
- Reduzir a confusão ao realizar processamento analítico de negócios.

Por fim, a consolidação dos datamarts conduz a processamentos econômica e operacionalmente eficientes.



-
- Reduzir a confusão ao realizar processamento analítico de negócios.

OPERANDO A PARTIR DE HUB CENTRALIZADO

Assim que os datamarts são consolidados, é possível, então, executar a tomada de decisões a partir de um hub centralizado, tal como o Hub de Distribuição de Dados, habilitado no System z. A essência do hub centralizado é que os dados detalhados são transacionados e coletados no hub e, dessa forma, são úteis para muitos departamentos da organização. E, assim que os dados são coletados no hub, podem, então, prover os datamarts consolidados.

Existem MUITAS vantagens em relação ao hub centralizado e datamarts consolidados. Algumas dessas vantagens são:

- Baixo custo operacional. Não há organizações periféricas trabalhando de maneira independente com dados de valor questionável.
- Integridade de dados. Com um hub centralizado, não há dúvida quanto ao local para onde se dirigir de forma a resolver um conflito de opinião.
- Realização de novas análises. Quando chega o momento de realizar uma nova análise, há uma fonte confiável e autorizada para a qual se dirigir.
- Eficiência operacional. Não apenas o custo de infraestrutura é menor quando há um hub centralizado, mas também o custo operacional de muitos programas e transações existentes no hub é reduzido.

Há, então, vantagens imediatas e vantagens estratégicas a longo prazo ao consolidar os datamarts, e operá-los a partir de um hub centralizado.

Nem é preciso ressaltar que organizações como a IBM possuem anos de experiência operacional com ambientes de hub centralizado. Esta experiência destaca-se na sofisticação da arquitetura e riqueza da infraestrutura que é suportada pela IBM e pelo System z.

Hub de Distribuição de Dados

Vamos analisar detalhadamente o hub de distribuição de dados. Trata-se de um tratamento das informações para todos os dados que são necessários aos datamarts. Todas as fontes são identificadas e os dados de todas as fontes são extraídos de uma única vez, em um período predeterminado. Então, todos os mercados são carregados com um conjunto de dados consistentes para análise.

Isso também proporciona insights sobre as redundâncias dos datamarts e, posteriormente, oportunidades para consolidá-los. O resultado são informações de um datamart, semelhantes a de outro datamart, com resultados confiáveis e que podem ser repetidos.

Reduzir Custos e Complexidade

Consolidar esses datamarts e o hub de distribuição de dados sobre um System z pode proporcionar todas as vantagens dos datamarts individuais, porém, sem grande parte de suas desvantagens. Utilizando a virtualização, cada datamart ainda pode ter sua individualidade e sua pequena comunidade de usuários. A IBM acredita que aproximadamente 20 processadores de datamarts possam ser consolidados em um único processador System z. Considerando que o software é, geralmente, licenciado pelo processador, isso resultaria em uma economia de 19:1. Em seguida, eliminam-se os servidores como mercadorias e, com eles, seu espaço físico e custos ambientais, como energia elétrica e refrigeração. E também são removidas todas as placas da interface de rede e a complexa rede que vincula todos os usuários a cada servidor de datamart, eliminando centenas de pontos de invasão de segurança.

Virtualização

Com os recursos de virtualização do System z, os novos datamarts são mais fáceis e rápidos de implementar, e geralmente, essa implementação ocorre em questão de horas, ao invés de dias ou semanas. Com um único ambiente administrativo cada mercado pode ser monitorado e administrado de modo simples e a partir de uma interface consistente. Quando os dados são extraídos do sistema operacional, podem ser rapidamente movidos para uma nova instância de banco de dados via HiperSockets, a uma velocidade de memória, para o Hub de Distribuição de Dados. Estes dados reorganizados podem então ser movidos novamente ao datamart, para utilização, sem despesas de deslocamento para fora do sistema. Para os dados que exigem alto grau de precisão, o software CDC pode sincronizar um datamart diretamente com o sistema de produção ou com o hub de distribuição de dados, o que poderia, então, sincronizá-lo com todos os recebimentos de dados dos datamarts. Isso fornece, quase que em tempo real, o acesso aos dados de produção.

Escapar da Armadilha?

Os datamarts podem proporcionar valor real, a um custo razoável, se monitorados e gerenciados de forma eficaz. Como qualquer recurso não gerenciado, podem se tornar um vírus fora de controle que busca consumir todos os recursos da organização. Mas, gerenciados de modo eficaz, podem ser um recurso que fornece informações com base em fatos, com uma abordagem consistente, oportuna e de custo eficaz.

Bill Inmon é conhecido como o “pai” do data warehouse. Bill escreveu 52 livros, traduzidos para nove idiomas. Um de seus livros vendeu mais de 500 mil cópias em todo o mundo. Em 2007, Bill foi nomeado pela ComputerWorld como uma das 10 pessoas mais influentes na história da computação. Seu empreendimento mais recente é o “textual disambiguation”, no qual dados textuais brutos são lidos e transformados em um formato de banco de dados relacional padrão.