

Alternativas de Integração de Dados - Gerenciando Valor e Qualidade

Utilizando uma abordagem regulada para incorporar serviços de qualidade de dados dentro do processo de integração de dados

WHITE PAPER:

QUALIDADE E INTEGRAÇÃO DE DADOS



Alternativas de Integração de Dados - Gerenciando Valor e Qualidade

Utilizando uma abordagem regulada para incorporar serviços de qualidade de dados dentro do processo de integração de dados

2

RESUMO

COM O ÍNDICE DE CRESCIMENTO DO VOLUME DE DADOS AUMENTANDO A UM RITMO ALUCINANTE, AS ORGANIZAÇÕES INTELIGENTES ESTÃO CADA VEZ MAIS CONTANDO COM RELATÓRIOS E ANÁLISES NÃO APENAS PARA CONDUZIR OS NEGÓCIOS, MAS TAMBÉM PARA MELHORAR A MANEIRA COMO OS NEGÓCIOS SÃO REALIZADOS. NO ENTANTO, EMBORA DÉCADAS DE INVESTIMENTOS EM APLICATIVOS EMPRESARIAIS TRANSACIONAIS E OPERACIONAIS TENHAM LEVADO A "ILHAS DE DADOS" VIRTUAIS, EXISTE UMA LISTA CRESCENTE DE APLICATIVOS EMPRESARIAIS COMO DE INTELIGÊNCIA DE NEGÓCIOS E ARMAZENAMENTO DE DADOS, GERENCIAMENTO DE RELACIONAMENTO COM O CLIENTE, PLANEJAMENTO DE RECURSOS EMPRESARIAIS E ABORDAGENS AINDA MAIS COMPLEXAS PARA ANÁLISES DE NEGÓCIOS QUE REQUEREM ACESSO A CONJUNTOS DE DADOS A PARTIR DE UMA SÉRIE DE ORIGENS.

OS PRINCIPAIS PROCESSOS DE GOVERNANÇA DE DADOS FACILITAM A COLETA E PADRONIZAÇÃO DOS REQUISITOS DE QUALIDADE DE DADOS CORPORATIVOS

A integração de dados está por toda a parte

Com o índice de crescimento do volume de dados aumentando a um ritmo alucinante, as organizações inteligentes estão cada vez mais contando com relatórios e análises não apenas para conduzir os negócios, mas também para melhorar a maneira como os negócios são realizados. A centralização de dados torna-se essencial para a implementação de aplicativos empresariais estratégicos. Os armazenamentos de dados operacionais, data warehouses, data marts, mashups, sistemas operacionais federados, relatórios de autoatendimento, trocas de dados e outros aplicativos analíticos e operacionais exigem um grau de compartilhamento de dados maior do que nunca. Atender às demandas por informação desses aplicativos empresariais de uso secundário torna-se o objetivo principal, e isso significa mover dados de fontes originais para sistemas de dados de negócios de destino. Em outras palavras, essas necessidades devem ser atendidas utilizando-se a integração de dados.

Alternativas de Integração de Dados

Conforme as necessidades dos consumidores de recebimento de dados se tornam mais sofisticadas, abordagens diferentes para a integração de dados evoluem. A abordagem mais tradicional “*Extract, Transform, Load*” (ETL) que leva os dados de suas origens para uma área temporária na qual os conjuntos de dados são manipulados e transformados em uma representação de destino. Uma abordagem alternativa é a virtualização de dados, na qual os dados permanecem armazenados na origem e uma visão conceitual é materializada sob demanda.

ETL tradicional

A abordagem mais comum para a integração de dados baseia-se em algumas variações do paradigma ETL. Como as origens de dados que são usadas para preencher aplicativos empresariais de uso secundário e de recebimento de dados muitas vezes estão em vários formatos diferentes, estruturas de arquivo e/ou de tabela,

e às vezes até mesmo usando codificações diferentes e subjacentes de caracteres, existe uma predisposição para normalizar representações de conjunto de dados antes de tentar combiná-las com um sistema de recebimento de dados de destino. O componente de extração implica rotinas especialmente projetadas empregadas para buscar dados das origens, que também irá exigir transformações especialmente projetadas que aplicam uma série de funções para normalizar, purificar, padronizar, derivar e converter, e outras funções necessárias para manipular os dados para um formato que seja adequadamente consistente com outras origens de dados transformadas em preparação para os sistemas de dados de destino. Nesse ponto, os dados estão prontos para serem propagados e carregados para o destino pretendido, substituindo os dados existentes ou periodicamente aumentando o conjunto de dados de destino existente.

Virtualização de dados

Ao contrário da abordagem tradicional de extração de dados de várias origens e de armazenamento temporário desses conjuntos de dados em uma área temporária, uma abordagem diferente, denominada federação de dados ou virtualização de dados, permite que os conjuntos de dados de origem permaneçam em seus locais originais. A virtualização de dados introduz camadas de abstração sobre uma série de origens de dados nativas e, como subproduto, proporciona vistas relacionais sem exigir que os dados sejam extraídos de sua origem. Essa abordagem para a abstração permite a criação de serviços de dados reutilizáveis, e as camadas de abstração de dados normalmente implementadas dentro de um ambiente de virtualização de dados permitem a apresentação de uma representação padronizada e lógica de conceitos de dados corporativos, permitindo assim que diversos consumidores de recebimento de dados tenham uma visualização dos dados estruturalmente e semanticamente consistente.

Alternativas de Integração de Dados - Gerenciando Valor e Qualidade

Utilizar uma abordagem regulada para incorporar serviços de qualidade de dados dentro do processo de integração de dados

4

Desafios relacionados a dados: integridade, consistência, razoabilidade

Com um interesse crescente no desenvolvimento de aplicativos empresariais que redirecionam as origens de dados primários para uso secundário, os consumidores de recebimento de dados podem ter expectativas muito diferentes com relação aos dados, principalmente em termos de qualidade de dados. Mas quando os conjuntos de dados são utilizados para fins aos quais não foram originalmente destinados, esses usuários de dados são muitas vezes forçados a redefinir e reinterpretar o significado dos conjuntos de dados originais.

Extrair e transformar os dados diversas vezes de diferentes maneiras para vários aplicativos pode levar a resultados variantes e à necessidade contínua de reconciliações, levando à desconfiança, tempo perdido, retrabalho e resultados questionáveis. Alguns problemas comuns incluem:

- A perda de valores de elementos de dados que distorcem dados e outras agregações;
- A variação no uso de dados de referência geralmente aceitos leva a inconsistências e imprecisões;
- As diferenças em formatos, estruturas e semântica presumidas por aplicativos empresariais de recebimento de dados podem levar a diferentes conclusões a partir de análises semelhantes;
- Inconsistência em relatórios que leva a uma necessidade contínua de reconciliação de relatórios gerados;
- Diferentes semânticas implícitas levam a interpretações equivocadas e a expectativas de razoabilidade perdidas.

Essencialmente, a ausência de padrões para estrutura, formatos e definições de dados redirecionados leva a problemas que surgem como resultado de processos de integração de dados não regulados - dados incompletos, dados inconsistentes e dados que não atendem às expectativas de razoabilidade.

Agregando valor utilizando serviços de qualidade de dados regulados

Se os desafios são introduzidos como um subproduto da necessidade de integração de dados, uma abordagem para enfrentá-los é reconhecer que os problemas existem, descobrir quais são os problemas e então reorganizar o processo de integração de dados para que incorpore meios de remediar falhas de dados antes que causem impacto relevante. Felizmente, uma abordagem regulada para incorporar serviços de qualidade de dados dentro do processo de integração de dados pode aliviar uma série de problemas que podem surgir como subproduto da consolidação desregulada de dados. O primeiro passo é instituir alguns processos essenciais de governança de dados para facilitar a coleta e padronização dos requisitos de qualidade de dados corporativos. O segundo passo é incorporar técnicas de gerenciamento de qualidade de dados dentro da estratégia de integração de dados.

Práticas de governança de dados

Embora um programa de governança de dados corporativos inclua uma grande variedade de processos de gerenciamento de dados, as exigências práticas da integração de dados sugerem o foco sobre um subconjunto dessas práticas que diretamente apoiam a integração de dados, como:

- Análise de requisitos de dados - O desenvolvimento típico de aplicativos empresariais considera a coleta de requisitos de dados como secundário ao processo de análise de requisitos funcionais. Mas como os projetos empresariais, como armazenamento de dados e gerenciamento de relacionamento com o cliente, cruzam os limites dos ramos de negócios, há uma necessidade de um processo bem definido para solicitar, documentar e sintetizar as expectativas de informações coletadas que todos os usuários de recebimento de dados terão, e então converter essas expectativas em requisitos de dados a serem impostos a todas as origens de dados candidatas. Isso não apenas impõe uma mudança radical na coleta de requisitos, também exige o tipo de supervisão fornecido por uma infraestrutura de governança de dados.

TENDO PRÁTICAS DE GOVERNANÇA DE DADOS SIMPLIFICA A INCORPORAÇÃO DAS TÉCNICAS DE QUALIDADE DE DADOS

- Revisão de padrões de dados - Definir padrões de dados pode resolver a questão da inconsistência, principalmente alinhando as semânticas e definições de elementos de dados. Quando as principais partes interessadas de toda a empresa participam de um processo de revisão e aprovação dos padrões de dados propostos, existe uma confiança de que os padrões serão definidos de forma que os requisitos coletados dos consumidores de recebimento de dados sejam atendidos.
- Gerenciamento de metadados - Inclui processos para documentar as definições e estruturas de padrões aprovados para domínios de dados de referência e troca de dados e fornecer um meio de comunicar esses padrões.

Serviços de qualidade de dados

Ter práticas de governança de dados simplifica a incorporação das técnicas de qualidade de dados, tais como:

- Análise e padronização - A análise é um processo que depende de formatos, padrões e estruturas definidas para determinar quando os valores de dados estão em conformidade com um conjunto comum de expectativas. A análise é usada com um conjunto de regras de padronização acionadas para transformar os dados de entrada em um formulário que possa ser usado de forma mais eficaz para padronizar a representação (presumindo uma representação válida) ou para corrigir os valores (caso erros conhecidos sejam identificados). A análise e padronização podem empregar uma biblioteca de domínios de dados e regras para dividir os valores de dados em vários componentes e reorganizar os componentes em um formato padronizado. A padronização também pode mudar palavras para abreviaturas ou abreviaturas para palavras, transformar apelidos em nome padrão, traduzir idiomas (por exemplo, espanhol para inglês), corrigir erros ortográficos comuns e reduzir a variação de valor para melhorar a associação de registros para deduplicação e limpeza de dados.
- Limpeza de dados - Quando os valores de dados são reconhecidos como sendo inconsistentes ou incorretos e as falhas dos dados não podem ser corrigidas no ponto de origem, uma alternativa é aplicar regras de transformação para imputar valores de dados, nomes ou endereços corretos, eliminar dados irrelevantes e/ou sem sentido, e até mesmo mesclar registros duplicados. A limpeza de dados diretamente garante que os dados que atendem a um nível de adequação, incorporando a mesma abordagem para análise, padronização e limpeza como parte do processo de integração de dados padronizem as transformações para que uma visão consistente dos dados seja fornecida a todos os consumidores de recebimento de dados.
- Validação de dados - Quando não há coordenação entre os consumidores de dados, eles podem não aplicar necessariamente as mesmas validações de dados da mesma forma. Mesmo se as regras fossem as mesmas ou semelhantes, quando as validações são aplicadas no ponto de uso, é improvável que as regras sejam aplicadas na mesma ordem, ou que os limites de aceitabilidade sejam os mesmos. Como resultado, embora as mesmas origens estejam sendo usadas, os resultados da validação também podem variar. Ao incorporar um conjunto padrão de validações de dados ao processo de integração de dados, as restrições podem ser testadas em pontos específicos no processo de fluxo de informações, reduzindo assim o risco de inconsistência. Solicitar requisitos de qualidade de dados da coleta de consumidores de recebimento de dados permite definir regras de qualidade de dados; implementar validação de conformidade com essas regras no início do processo pode ajudar a garantir que a qualidade dos dados seja suficiente para atender às necessidades dos negócios e permitir que possíveis problemas sejam identificados e corrigidos antecipadamente e de forma consistente.

Alternativas de Integração de Dados - Gerenciando Valor e Qualidade

Utilizando uma abordagem regulada para incorporar serviços de qualidade de dados dentro do processo de integração de dados

6

Considerações

A abordagem utilizada para implementar a integração de dados não deve interferir no desejo de melhorar a qualidade dos dados de forma coerente e consistente. A maioria dos fornecedores de ferramentas ETL tradicionais reconheceu a necessidade de incorporar técnicas de qualidade de dados, e muitos formaram relações de parceria ou conquistaram fornecedores de ferramentas de qualidade de dados. Hoje, é raro encontrar uma ferramenta de extração de ponta a ponta, de transformação e de carregamento que não estimula a definição e implementação de análise, padronização, limpeza e validação incorporadas.

Em vez disso, as ferramentas de virtualização de dados estão também cada vez mais presas a ferramentas de qualidade de dados e à tecnologia. A abstração fornecida através da virtualização de dados oferece uma oportunidade para a equipe de gerenciamento de dados envolver os consumidores de recebimento de dados, solicitar seus requisitos de qualidade de dados e diretamente integrar validações baseadas em atributos de dados dentro de uma das camadas de abstração. Consolidar requisitos de qualidade de dados e implementar restrições de validação de dados em pontos específicos no fluxo de produção de informações reduz o risco de inconsistência, ajuda a garantir que a qualidade dos dados seja suficiente para atender às necessidades de consumo de recebimento de dados, e alerta os administradores de dados quanto a possíveis problemas que podem ser identificados e remediados no início do processo de integração de dados.

O resultado: a integração de dados está se tornando dominante em toda a organização. Com a introdução de processo regulado que simplifica a reutilização de informações de forma consistente, a confiança nos relatórios e análises aumentará, beneficiando todos os interessados em toda a organização!

PARA OBTER MAIS INFORMAÇÕES SOBRE SOLUÇÕES DE QUALIDADE E INTEGRAÇÃO DE DADOS, LIGUE PARA A PITNEY BOWES SOFTWARE HOJE OU ACESSE NOSSO SITE.

David Loshin é o Presidente da Knowledge Integrity, Inc., uma empresa de consultoria e desenvolvimento com foco em soluções personalizadas de gerenciamento de informações, incluindo consultoria de soluções de qualidade de informações, treinamento de qualidade de informações e soluções de regras de negócios. Loshin é o autor dos livros *Master Data Management*, *Enterprise Knowledge Management – The Data Quality Approach* e *Business Intelligence – The Savvy Manager’s Guide* e realiza palestras sobre valorização das informações. Entre em contato com David enviando e-mail para loshin@knowledge-integrity.com ou ligando para o número (301) 754-6350.

www.pb.com/software

ESTADOS UNIDOS

One Global View
Troy, NY 12180
telefone: +1.518.285.6000
fax: +1.518.285.6070
1.800.327.8627
la.sales@pb.com
www.pbinsight.com

ESCRITÓRIO MIAMI

15100 NW 67th Avenue, Suite 403
Miami Lakes, FL 33014
U.S.A.
telefone: +1.305.825.8381
fax: +1.305.826.1246
la.sales@pb.com
latam.pbinsight.com

ESCRITÓRIO MÉXICO

San Francisco 1393
Colonia Del Valle
Delegación Benito Juárez
03210 México, DF
México
telefone: +52.55.5636.0949
la.sales@pb.com
www.pbinsight.com.mx

ESCRITÓRIO BRASIL

Av. Engenheiro Luis Carlos Berrini,
1.681- 9º andar
04671-903 São Paulo, SP
Brasil
telefone: +55 11 5698 5570
la.sales@pb.com
www.pbinsight.com.br