

Integração de Dados I

Definições, Requisitos, Características, Razões de Desafios

Prof. Marco Paulo Soares Gomes

Bibliografia

- material do Prof. Claudiney Ramos

Integração de Dados: Definições

- A Integração de Dados é o processo de combinar dados de múltiplas fontes e em diferentes formatos e armazená-los em um repositório de destino, de forma que os dados fiquem acessíveis para as aplicações de tomada de decisão e/ou para projetos de dados.
- É um processo chave para transformar os dados brutos em *insights* que levam a uma tomada de decisão mais rápida e assertiva.

Integração x Replicação x Migração de Dados

- **Integração de dados:** é o processo de mover diferentes conjuntos de dados de bancos de dados, plataformas de nuvem, e outras fontes para um repositório de dados centralizado.
- **Replicação de dados:** é o armazenamento de cópias dos dados em diferentes locais, geralmente para melhorar a disponibilidade de dados ou a resiliência do sistema.
- **Migração de dados:** é um evento em que os dados são movimentados (movidos) de uma plataforma para outra, e a plataforma de destino torna-se a nova fonte dos dados.

Integração de Dados: Definições

- **Fontes de dados (ou *data sources*):**

são os locais onde os dados estão armazenados, contendo dados úteis que podem contribuir para a tomada de decisão e gerar algum tipo de inteligência.

Integração de Dados: Definições

- **Tipos de fontes de dados:**
 - Arquivos de texto
 - Planilhas eletrônicas
 - Conexões com bancos de dados
 - Arquivos em formato CSV, JSON e XML, por exemplo
 - APIs de integração de dados

Integração de Dados: Requisitos

- A identificação dos requisitos de integração de dados inclui:
 - Entendimento das fontes de dados;
 - Mapeamento dos sistemas informatizados, que incluem os bancos de dados transacionais (OLTP) que mantêm os dados gerados e manipulados pelas aplicações do negócio;
 - Mapeamento de Data Warehouse (DW), Data Mart (DM) e Data Lakes.

Integração de Dados: Requisitos

- A identificação dos requisitos de integração de dados inclui (continuação):
 - Levantamento de outras fontes de dados: planilhas, documentos de texto, e-mails, e outras fontes de dados internas e externas, como as redes sociais, por exemplo;
 - Definir a periodicidade da extração dos dados, ou se o processo será executado de forma permanente (on-line);
 - No contexto da integração em larga escala (*big data*), é necessário dimensionar o sistema para suportar o grande volume de dados.

Integração de Dados: Requisitos

- A identificação dos requisitos de integração de dados inclui (continuação):
 - Monitorar constantemente o surgimento de novas necessidades para os negócios e novos requisitos de integração que podem surgir.

Integração de Dados: Características

- uso de processos técnicos e de negócios para combinar dados de diversas fontes, com o objetivo de fornecer informação útil e valiosa, de forma eficiente.
- colaboração entre usuários internos e externos à organização.
- os dados devem estar disponíveis e acessíveis a partir de uma localização central (*Data Mart, Data Warehouse* ou *Data Lake*).

Integração de Dados: Características

- a integração de dados deve ser um processo colaborativo e unificado que melhora a comunicação e a eficiência de uma organização, e a torna mais competitiva e produtiva.
- quando uma organização integra seus dados, o tempo para preparar e analisar os dados é reduzido significativamente.
- visões unificadas tornam-se automatizadas, e os departamentos não terão necessidade de construir novas conexões para criar relatórios ou construir novas aplicações.

Integração de Dados: Razões

- Dados integrados reduzem erros e retrabalho.
- A integração de dados melhora, significativamente, a qualidade dos dados de uma organização.
- A integração de dados permite que uma organização faça mais, usando menos recursos.

Integração de Dados: Desafios

- Unificar diversas fontes de dados e combiná-las em uma estrutura simples pode ser um processo bastante difícil.
- As organizações devem criar processos para mover e traduzir os dados que ainda não estão integrados.
- Fontes de dados externas: baixa qualidade dos dados, questões de privacidade dos dados.

Integração de Dados: Desafios

- Sistemas legados: exigem maior esforço devido às características destes sistemas.
- Sistemas de “ponta”: dados não estruturados e de tempo real (dispositivos IoT, sensores, vídeos, áudio e nuvem).
- Mapeamento das fontes de dados e dos tipos de análise (uso dos dados integrados).
- Evolução dos dados: manter os esforços de integração (novas demandas e mudança de necessidades da organização).

Integração de Dados: Métodos

- Integração de dados manual: é o processo de coletar manualmente os dados necessários de diversas fontes, os acessando diretamente.
- *Integração baseada em aplicação: usa aplicações de software para localizar, recuperar, e integrar os dados.*
 - *Ex.: middleware* de integração de dados: age como um mediador, e ajuda a normalizar os dados antes de trazê-los para o “pool” de dados mestres (repositório central).
- Integração de armazenamento comum: é a abordagem mais popular de integração de dados.
 - utiliza um repositório central (bancos de dados transacionais, Data Warehouse, ...) para armazenar os dados integrados (visão unificada).

Dúvidas, Perguntas ou Sugestões?