



Qualidade de Dados em Sistema de Integração

Ivson de Assis Cavalcanti

iac@cin.ufpe.br

Roteiro

1. Motivação
2. Sistema de Integração de Dados
 1. Arquitetura Data Warehouse
 2. Arquitetura Mediator
3. Qualidade de Informação (QI)
4. Critérios de avaliação de QI em Sistemas de Integração
 1. Critério subjetivo
 2. Critério objetivo
 3. Critério de processo
5. QI em Sistemas de Integração
6. Conclusão
7. Referencias

Motivação

- ▶ Enorme volume de dados
- ▶ Heterogeneidade –
 - ▶ nível físico (diferentes plataformas de hardware e software)
 - ▶ nível lógico (diferentes modelos de dados).
 - ▶ nível conceitual (diferentes esquemas e conceitos).

Sistema de Integração

- ▶ permitir que usuários consultem simultaneamente múltiplas fontes de dados
 - ▶ Heterogêneas
 - ▶ Distribuídas
 - ▶ Autônomas
- ▶ Manter transparentes os procedimentos de acesso, extração e integração dos dados

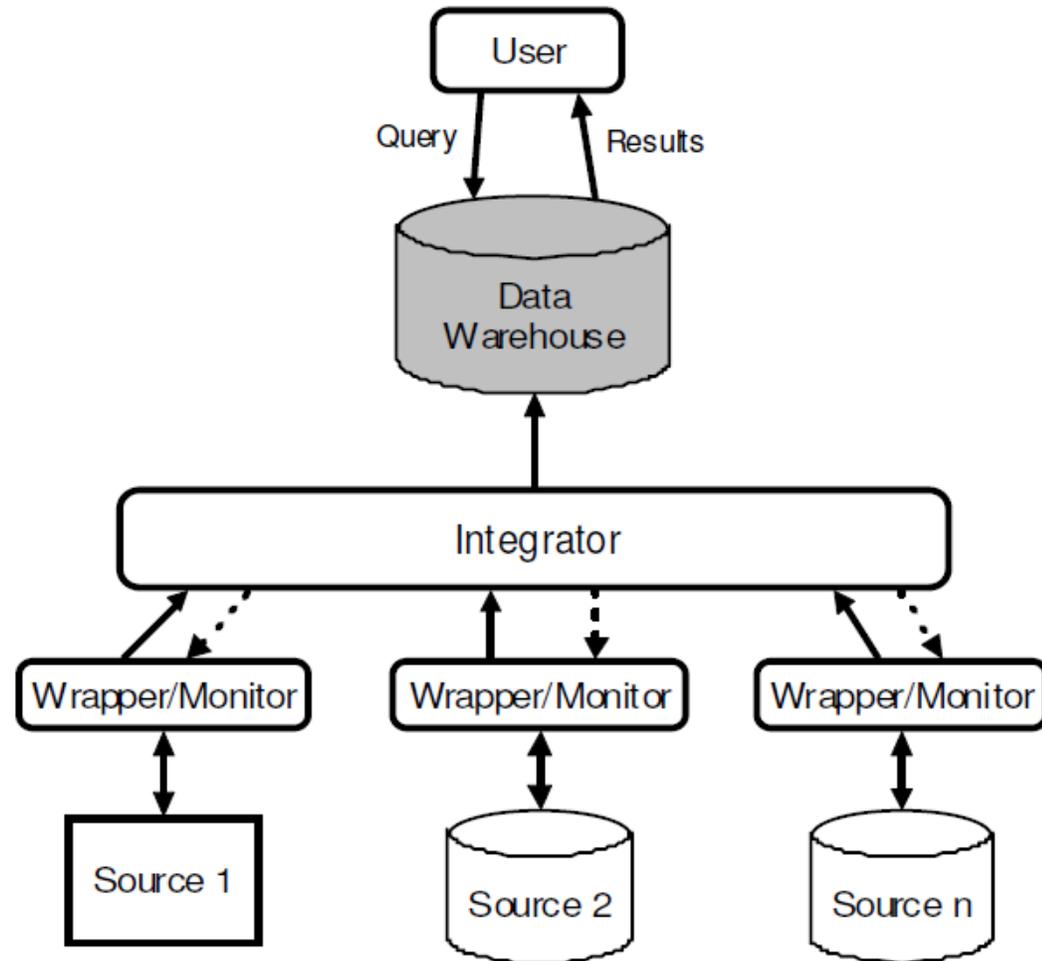
Sistema de Integração

- ▶ Abordagens
 - ▶ Materializada/Warehousing
 - ▶ Virtual/Mediador

Warehousing

- ▶ Os dados das diversas fontes são carregados (materializados) em um repositório (data warehouse) e as consultas são submetidas a este repositório.
- ▶ **Vantagem: performance garantida no momento da consulta**
- ▶ **Desvantagem: atualização do repositório sempre que houver mudança nos dados**

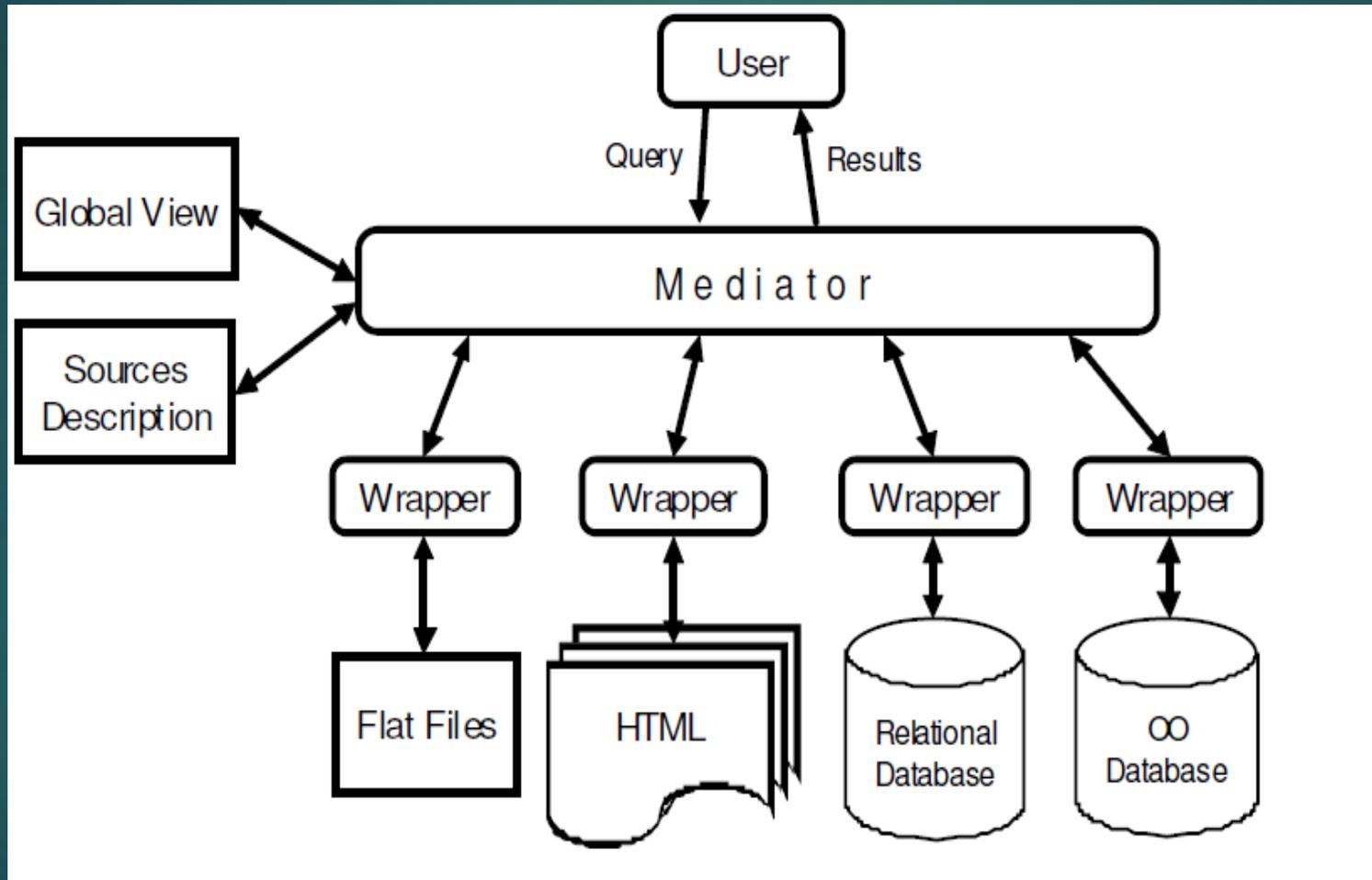
Arquitetura Materializada



Virtualização

- ▶ os dados são mantidos nas fontes e as consultas são submetidas diretamente às diversas fontes
- ▶ **Vantagem:** os dados não são replicados e tem-se a garantia de estarem atualizados no momento da consulta
- ▶ **Desvantagem:** como as fontes de dados são autônomas, são necessários métodos para otimização de consultas para garantir uma performance adequada

Arquitetura Mediador



Qualidade da Informação

- ▶ Qualidade é Subjetivo
- ▶ Qualidade da Informação
 - ▶ Informações de boa qualidade são informações que podem ser “usadas pelos consumidores” Fitness for use
- ▶ A falta de qualidade nas informações pode impactar severamente a efetividade de uma organização O crescimento do uso de tecnologias de data warehouse e acesso direto a fontes de dados distribuídas
- ▶ tornou-se um ponto crítico em empresas e em pesquisas Ainda mais com o volume/velocidade/ variedade



Critério de Avaliação em Sistema de Integração

- ▶ Richard Wang e Naumann
- ▶ Critérios para Avaliação de Qualidade da informação
- ▶ Fatores principais
 - ▶ percepção do usuário
 - ▶ informação em si
 - ▶ processo para acesso a informação



Critério subjetivo

- ▶ Percepção do usuário
 - ▶ Credibilidade
 - ▶ Representação Concisa
 - ▶ Interpretabilidade
 - ▶ Relevância
 - ▶ Reputação
 - ▶ Compreensibilidade
 - ▶ Valor Agregado

Classe	Critério de QI	Método de Avaliação
Critérios de Sujeito	Credibilidade	Experiência do Usuário
	Representação Concisa	Entradas do Usuário
	Interpretabilidade	Entradas do Usuário
	Relevância	Avaliação Contínua
	Reputação	Experiência do Usuário
	Compreensibilidade	Entradas do Usuário
	Valor agregado	Avaliação Contínua

Critério objetivo

- ▶ A informação em si
 - ▶ Completude
 - ▶ Suporte ao Usuário
 - ▶ Documentação
 - ▶ Preço
 - ▶ Confiabilidade
 - ▶ Segurança
 - ▶ Atualidade
 - ▶ Verificabilidade

Classe	Critério de QI	Método de Avaliação
Critérios de Objeto	Completude	Parsing, Entradas do Usuário
	Suporte ao Usuário	Parsing, Contrato
	Documentação	Parsing
	Objetividade	Entradas de Especialista
	Preço	Contrato
	Confiabilidade	Avaliação Contínua
	Segurança	Parsing
	Atualidade	Parsing
	Verificabilidade	Entradas de Especialista

Critério de Processo

- ▶ Processo de integração em si
 - ▶ Precisão
 - ▶ Volume de Dados
 - ▶ Disponibilidade
 - ▶ Latência
 - ▶ Tempo de Resposta
 - ▶ Representação Consistente

Classe	Critério de QI	Método de Avaliação
Critérios de Processo	Precisão	Avaliação Contínua
	Volume de Dados	Avaliação Contínua
	Disponibilidade	Avaliação Contínua
	Latência	Avaliação Contínua
	Tempo de Resposta	Avaliação Contínua
	Representação Consistente	Avaliação Contínua

QI em Sistemas de Integração

Elemento do Sistema	Crítérios relevantes QI
Data source Shema	Reputação, Completude, Disponibilidade, Atualidade
Mediation Shema	Completude, Minimalismo
Mediation Queries	Completude, Disponibilidade, Atualidade, Tempo de Resposta, Minimalismo
Source Seleciton	Reputação, Completude, Atualidade, Verificabilidade
Query Process	Disponibilidade, Tempo de Resposta, Frequência de Acesso
Data Materialization	Atualidade, Tempo de Resposta, Frequência de Acesso, Disponibilidade, Reputação, Verificabilidade
Data Integration	Completude, Atualidade, Verificabilidade, Precisão

Conclusão

- ▶ A quantidade de informação disponível vem sempre crescendo. E de maneira distribuída, autônoma e dinâmica. Consequentemente vem a necessidade de sistemas de integração para obter informação de forma eficiente e com qualidade. Através do estudo apresentado vimos como Avaliar sistemas de integração de maneira generalizada através de diversos critérios de qualidade de informação.

Referências

- ▶ [1] Maria C. M. Batista, 2006. “*Information Quality Analysis in a Data Integration System*”
- ▶ [2] JOSKO, João Marcelo B. 19 maio 2008: “[Uma análise das alternativas de arquiteturas para a integração de dados](#)”, Acessado em 22 de setembro de 2016
- ▶ [3] Erika Maria 19 outubro 2007: “[Uma visão sobre a Qualidade dos Dados](#)”, Acessado em 22 de setembro de 2016
- ▶ [4] Edson C. Germano e Hiroo T., 2012. “[Uma Análise das Dimensões da Qualidade de Dados em Projetos de Dados Governamentais Abertos](#)” Acessado em 22 de setembro de 2016
- ▶ [5] “[Integração de Dados](#)” Acessada em 22 de setembro de 2016
- ▶ [6] “[Análise de Critérios de Qualidade da Informação em um PDMS](#)” Acessada em 22 de setembro de 2016

