

ANÁLISE VISUAL DE BASE DE DADOS – UM ESTUDO DE CASO

Carlos Timóteo, Edgar Almeida e Maxwell Queiroz

INTRODUÇÃO

- ✖ Aplicações para análise de dados e relatórios enfrentam os seguintes problemas[1]:
 - + Apresentação de Informações de forma inefetiva;
 - + Falta de capacidade exploratória;
 - + Interfaces de Usuários difíceis.

O QUE SE ESPERA ?

- ✖ Interface da aplicação com alto desempenho para que questões que envolvem multi-dimensões possam ser respondidas;
- ✖ Ferramenta para análise exploratória de forma livre;
- ✖ Interface de análise fácil para trabalhadores do conhecimento;
- ✖ *Uma nova forma de interação com banco de dados e planilhas.*

OBJETIVOS

- ✖ Estudar o estado da arte em análise visual de base de dados;
- ✖ Estudo de caso, com a base de dados indexada dos trabalhos científicos publicados no Simpósio Brasileiro de Telecomunicação(SBrT)[8] ;
- ✖ Avaliação de desempenho entre as ferramentas *Tableau* e *Pentaho*;

REVISÃO DE LITERATURA

✖ Análise Visual:

- + Raciocínio analítico facilitado por interfaces visuais interativas[4].
- + Combina técnicas de análise automatizada com visualizações interativas.
- + Fornecer avaliações oportunas, defensáveis e compreensíveis. Prover maior capacidade de “exploração” e exploração.

REVISÃO DE LITERATURA

- ✖ **Extração – Transformação – Carga (ETL)[5]:**
 - + Processo de software que facilita o preenchimento de *data warehouses* – áreas para tratamento e processamento de dados.
 - + O resultado é observado de forma visual.

REVISÃO DE LITERATURA

× VizQL[6]:

- + Linguagens de consulta convencionais, tais como SQL e MDX têm capacidades de formatação e visualização limitadas. VizQL é projetada para preencher essa lacuna;
- + Linguagem declarativa. A imagem desejada é descrita, as operações de baixo nível são geradas automaticamente pelo analisador de consulta.
- + Casamento entre consulta, análise e visualização.

REVISÃO DE LITERATURA

✖ Business Intelligence (BI):

- + De acordo com [7], *Business Intelligence* (BI) é um conjunto de tecnologias que tem como objetivo prover e oferecer suporte a um ambiente de informação;
- + O problema de negócios atual deve ser quebrado em partes e segregado em tarefas relativas a Relatório, Análise, Mineração de Dados, *Workflow*, etc;

REVISÃO DE LITERATURA

× Pentaho[2]

- + Sistema de código aberto para BI, desenvolvido em linguagem Java;
- + A plataforma de BI fornece uma estrutura de execução e serviços que incluem registro, auditoria, segurança, programação, ETL, serviços web, repositório de atributo e motores de regras.

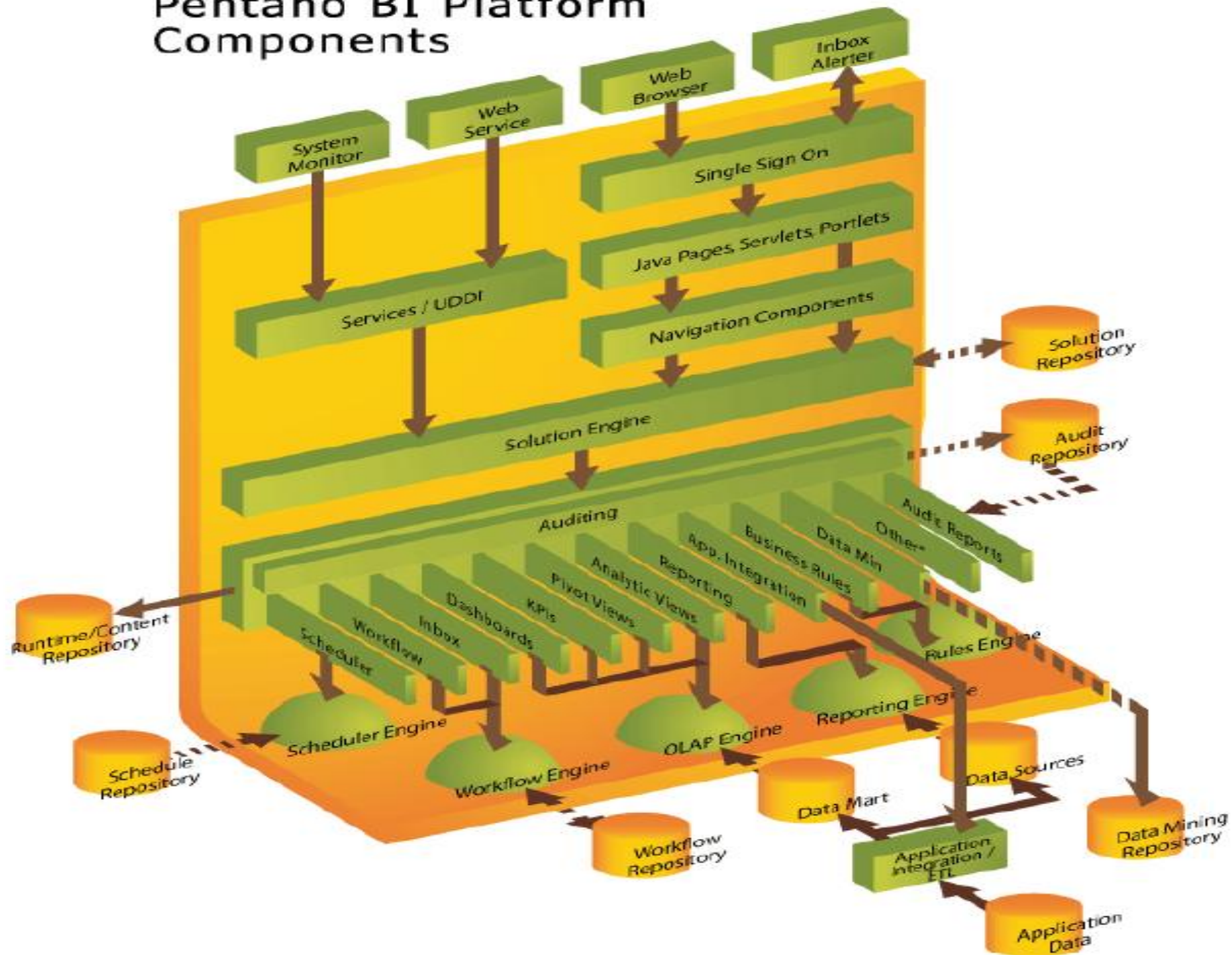
REVISÃO DE LITERATURA

✕ Pentaho

✕ The services of the BI platform:

- + • Provide web services to external applications;
- + • Have access to the same Solution Engine as the user interface components;
- + • Are called by the workflow engine and scheduler to execute system actions;

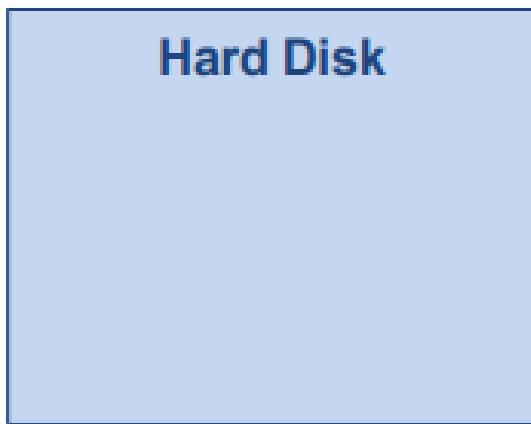
Pentaho BI Platform Components



REVISÃO DE LITERATURA

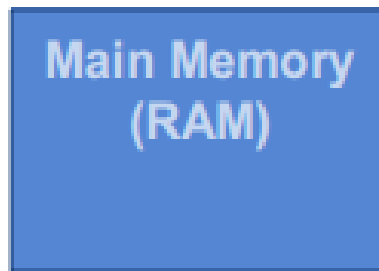
× Tableau[3]

- + A análise e visualização devem ser sinergicamente integradas em um processo de análise visual. A análise visual significa especificamente:
 - + Data Exploration
 - + Data Visualization



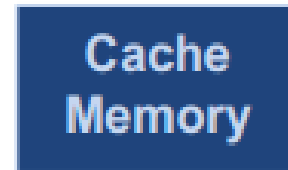
Hard Disk

Big, slow memory



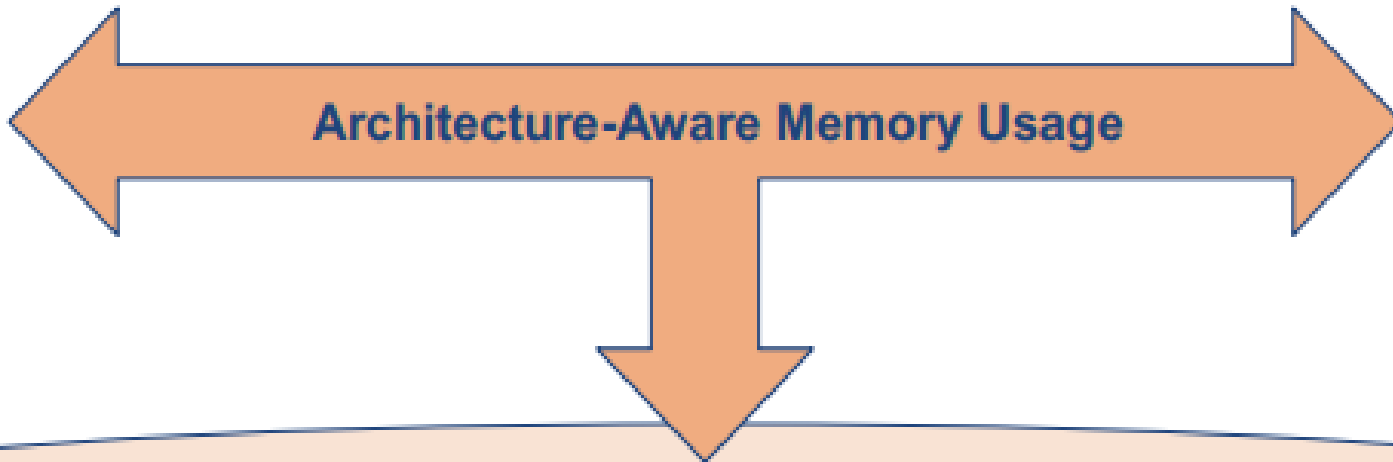
**Main Memory
(RAM)**

*Fast, small
memory*



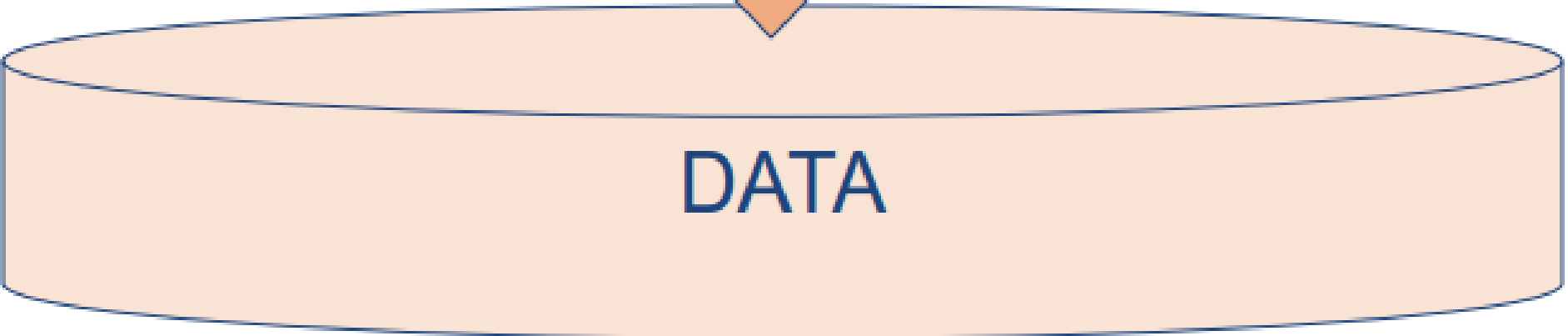
**Cache
Memory**

*Very fast,
very small
memory*



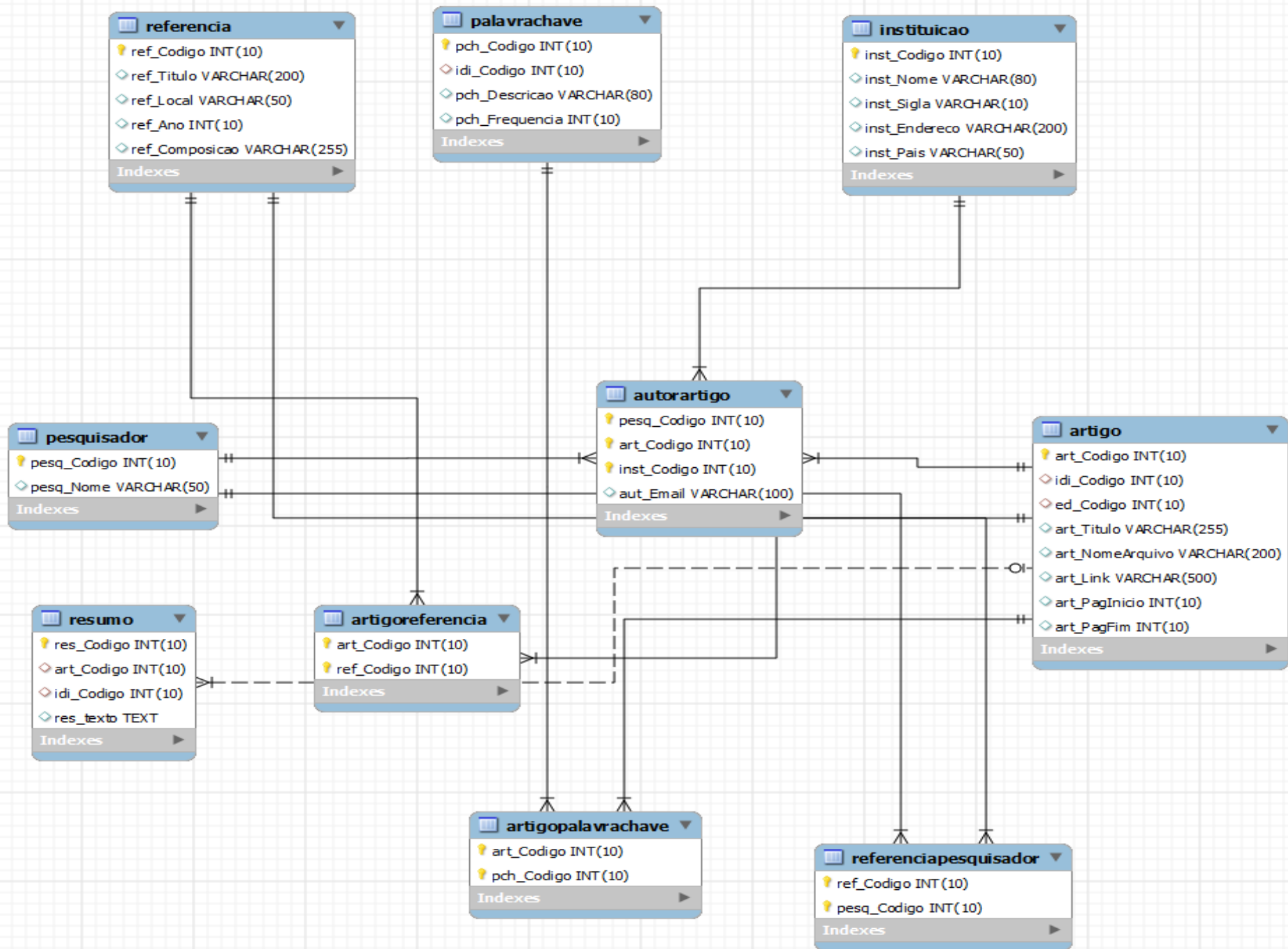
Architecture-Aware Memory Usage

DATA



ESTUDO DE CASO

- ✗ Base de dados com informações extraídas dos trabalhos científicos publicados.
- + Dificuldades:
 - ✗ Incompleto (schema);
 - ✗ Poucos registros;
 - ✗ Problemas na extração da informação;



ESTUDO DE CASO

× Tableau

+ Passos para Exploração de Dados:

- × (1) Filtragem para se focar em itens de interesse;
- × (2) Ordenação para classificar e priorizar;
- × (3) Agrupando e agregando para resumir;
- × (4) Criando cálculos em tempo real para expressar número em formas úteis;

ESTUDO DE CASO

✕ Perguntas

- + Qual autor mais publicou?
- + Qual referência da SBrT foi mais citada?
- + Qual referência de todos os eventos foi mais citada?
- + Quais as palavras-chaves mais citadas?

ESTUDO DE CASO

Prática!

DÚVIDAS?

FIM.

REFERÊNCIAS

- [1] Keim, D.A., Andrienko, G., Fekete, J.D., Gorg, C., Kohlhammer, J., Melançon, G: Visual Analytics: Definition, Process, and Challenges. A. Kerren et al. (Eds.): Information Visualization, LNCS 4950, pp. 154–175, 2008.
- [2] Pentaho (2012). Pentaho BI. <http://www.pentaho.com/> - Acessado em 02/12/2012.
- [3] Tableau (2012). Tableau Software. <http://www.tableausoftware.com/> - Acessado em 02/12/2012.
- [4] Thomas, J.J., Cook, K.A.: Illuminating the Path. IEEE Computer Society Press, Los Alamitos (2005).
- [5] Vassiliadis, P. A Survey of Extract-Transform-Load Technology. International Journal of Data Warehousing & Mining, 5(3), 1-27, July-September 2009.
- [6] Hanrahan, P. VizQL: A Language for Query, Analysis and Visualization. SIGMOD 2006, June 27-29, 2006, Chicago, Illinois, USA.
- [7] Petrini, M., Freitas, M. T., e Pozzebon, M. (2006). Inteligência de negócios ou inteligência competitiva: noivo neurótico, noiva nervosa. Encontro da Associação Nacional de Pós-Graduação e Pesquisa em Administração (EnANPAD).
- [8] Alves, N., Lencastre, M., Lins, R. Improving Requirements Quality in Digital Libraries: The case of Scientific Proceedings. In Proceedings of the Quatic 2012-8th International Conference on the Quality of Information and Communications Technology. Lisbon, Portugal, September 3-6, 2012.