



Modelagem Dimensional, Indicadores de Desempenho e Dashboards

Centro Universitário 7
Setembro - Uni7

**Especialização em
Ciência de Dados com Big
Data, BI e Data Analytics**

Prof. Dr. Manoel Ribeiro

Encontros

- Dia 1
 - Apresentação
 - Aula motivacional - Qual a importância de medir?
- Dia 2
 - Conceitos fundamentais de Data warehouse

Encontros

- Dia 3 (sexta)
 - Conceitos de indicador e media
- Dia 4 (sábado)
 - BSC e Dashboard
 - Exemplo de KPI para área de TIC
- Dia 5 (sexta)
 - Laboratório de modelagem multidimensional
- Dia 6 (sábado)
 - Avaliação

Repositório

<https://github.com/antoniomralmeida/mdi>

Modelagem DW

Princípios para modelagem DW

- . Modelo Star Scheme
- . Reduzir granularidade (Análise de granularidade)
- . Utilizar apenas campos numéricos na tabela fato
- . Carga do Stage por mecanismo de Delta
- . Operação de exclusão somente para dados recentes
- . Consistência da métrica com granularidade (Análise de consistência das métricas)

Cuidados na modelagem DW

- . Cuidado com integridade das dimensões (usar PK)
- . Proibido reaproveitamento de código em cadastro
- . Proibido mudança de unidade de medida ou fator de escala
- . Cuidado com atributos multivalorados de dimensões

Passo a passo para construção de uma estrela

- Identificação do conjunto de tabelas que formarão a tabela fato
 - construir o select com saída no formato da tabela fato, este select deverá conter o id da tabela de maior detalhe ou deve-se gerar um hash
- Identificar métricas que não estão na mesma granularidade da tabela fato
 - Identificar select que define o critério de rateio para esta métrica
- Identificação do conjunto de tabelas que comporão cada dimensão

Preparação do Delta no Stage

- Criar uma tabela de Delta com timestamp, id da tabela de maior detalhe, operação (1-incluir, 2-alterar, 3-excluir), status (1-pendente, 2-processando, 3-processado)
- Na tabela base de maior detalhe criar uma trigger para gravar as alterações na tabela de Delta
- Se houver tabela com métricas num nível acima da tabela de maior detalhe, esta deve ter uma trigger gravando as alterações na tabela de Delta referenciando os id da tabela de maior detalhe relacionados

Análise de granularidade

```
select 'base_select', count(*) from (base select) a  
union all
```

```
select 'tabela_fato', count(*) from (select tabela_fato) a
```

Análises:

1. base_select deve ser SEMPRE maior que tabela_fato
2. tabela_fato deve ser o menor possível

Análise de consistência das métricas

`select 'base_table', sum(metrica) from (base table) a
union all`

`select 'tabela_fato', sum(metrica) from (select tabela_fato) a`

Análises:

1. `base_table` deve ser SEMPRE igual a `tabela_fato`
2. Cada métrica deve ser analisada isoladamente
3. diferenças podem ser resultantes de relacionamentos opcionais, relações quebradas (bugs) ou produto cartesianos

Prática com database Sakila

Sakila

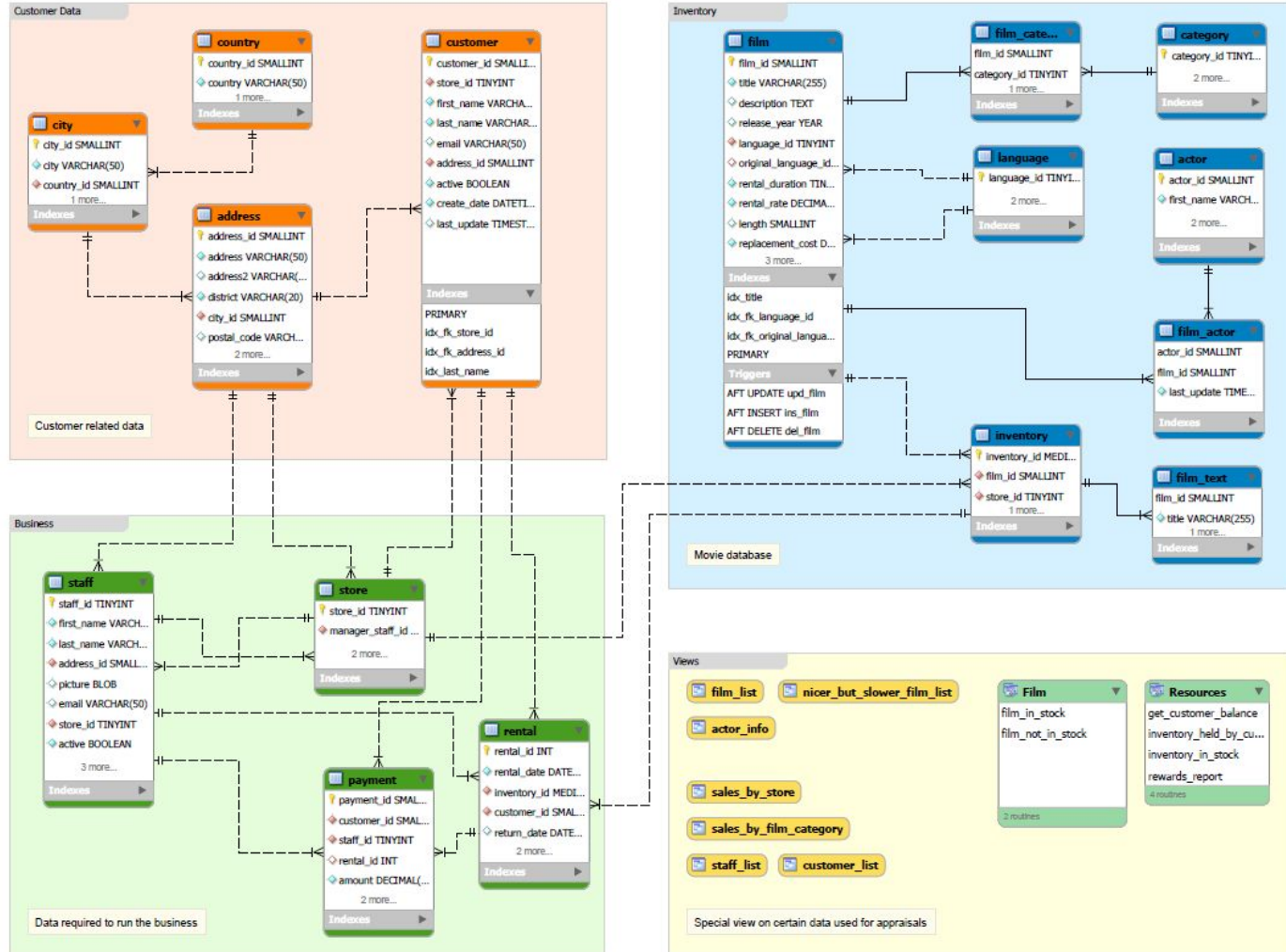
O banco de dados de amostras Sakila foi projetado como um substituto para o world banco de dados de amostra, também fornecido pela Oracle.

O `world` banco de dados de amostra fornece um conjunto de tabelas contendo informações sobre os países e cidades do mundo e é útil para consultas básicas, mas carece de estruturas para testar a funcionalidade específica do MySQL e os novos recursos encontrados no MySQL 5.

O desenvolvimento do banco de dados de amostras Sakila começou no início de 2005. Os primeiros projetos foram baseados no banco de dados usado no white paper da Dell, [Three Approaches to MySQL Applications, nos servidores Dell PowerEdge](#) .

Onde o banco de dados de amostra da Dell foi projetado para representar uma loja de DVDs on-line, o banco de dados de amostras Sakila foi projetado para representar uma loja de locação de DVDs. O banco de dados de amostras Sakila ainda empresta nomes de filmes e atores do banco de dados de amostra da Dell.

Sakila



Sakila

Construir um cubo OLAP que permita melhor gerenciar o aluguel de filmes, indicando os filmes e clientes mais rentáveis, bem como os clientes com mais atraso e os filmes com maior período de aluguel e filmes com maior taxa de retorno valor/período



Pentaho

Pentaho é um software de código aberto para inteligência empresarial, desenvolvido em Java.

A solução cobre as áreas de ETL (Extraction, Transformation and Load), reporting, OLAP e mineração de dados (data-mining).

Desenvolvido desde 2004 pela Pentaho Corporation o software foi considerado uma das melhores aplicações para inteligência empresarial em 2008 pela InfoWorld.

Existem as versões:

- PENTAHO BUSINESS ANALYTICS (licenciada)
- COMMUNITY EDITION (livre)

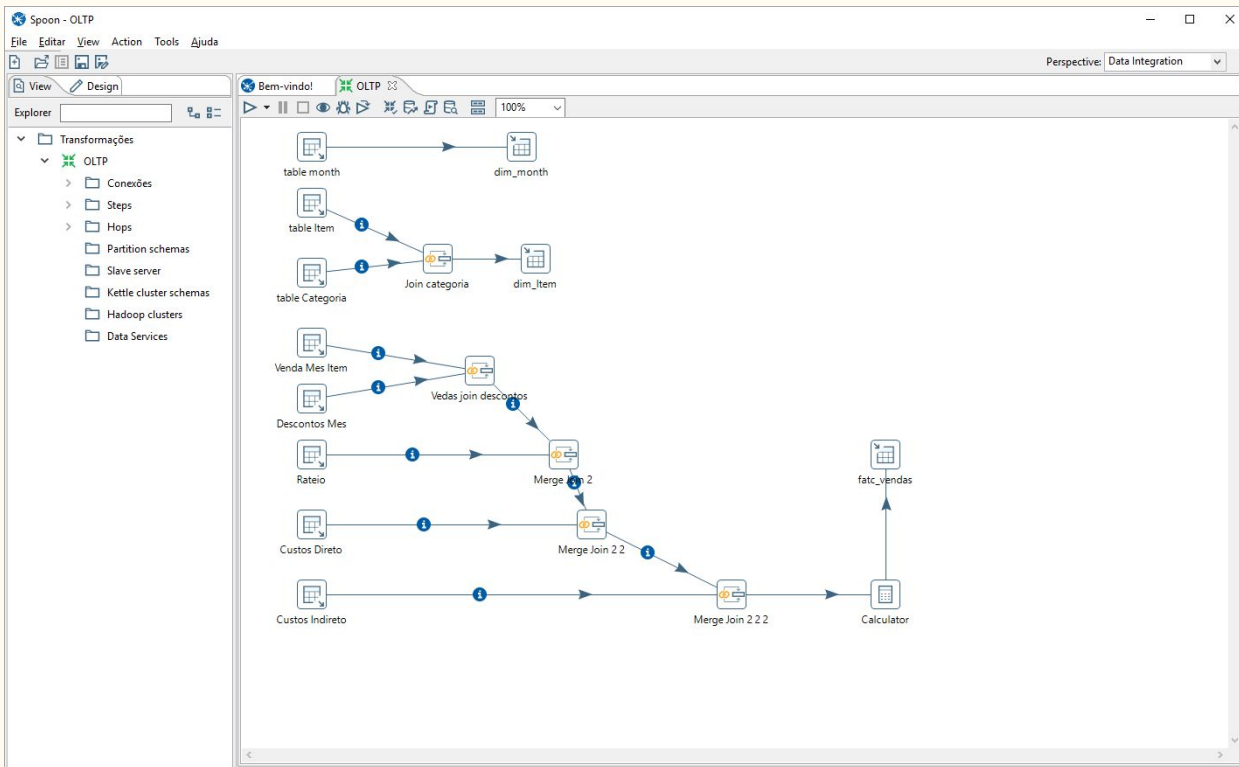
Pentaho Community Edition

<http://community.pentaho.com/>

Módulos:

- Data Integration (Spoon/Kitchen) - ETL
- Schema Workbench – OLAP
- BI Server – Jpivot, Dashboard, Mondrian
- Report Designer
- Marketplace (Add-On)

Data Integration - Spoon



Planejamento do ETL

ETL			
Dimensão/Fato/Mensuração	Origem	formula	SQL
dim_item	Item Categoria		<code>select i.item, i.descricao, c.descricao as categoria from item i join categoria c on i.idcategoria = c.idcategoria</code>
dim_month	Venda		<code>select distinct month(data) as month, fnt_date(month(data)) as fnt_month, halfyear(data) as halfyear, fnt_date(halfyear(data)) as fnt_halfyear, quarter(data) as quarter, fnt_date(quarter(data)) as fnt_quarter, year(data) as year from venda v</code>
DescontoRateio	A*FATOR	A	<code>select month(data) as month, sum(desconto) as desconto from venda v group by month(data)</code>
Fator		B/C	
		B	<code>select month(data) as month, i.item, sum(i.quantidade*valor_unitario) as vlbruto from venda v join item_venda i on i.idvenda = v.idvenda group by month(data), i.item</code>
		C	<code>select month(data) as month, sum(i.quantidade*valor_unitario) as base from venda v join item_venda i on i.idvenda = v.idvenda group by month(data)</code>
CustoDiretoRateio	D*FATOR	D	<code>select mes_ref as month, sum(valor_total) as custodireto from custos c join rubrica r on c.idrubrica = r.idrubrica where sigla = 'cdi' group by mes_ref</code>
CustoIndiretoRateio	E*FATOR	E	<code>select mes_ref as month, sum(valor_total) as custodireto from custos c join rubrica r on c.idrubrica = r.idrubrica where sigla <> 'cdi' group by mes_ref</code>

Schema Workbench

Schema Workbench

File Edit View Options Windows Help

Schema - FAT (Vendas.xml)

Schema

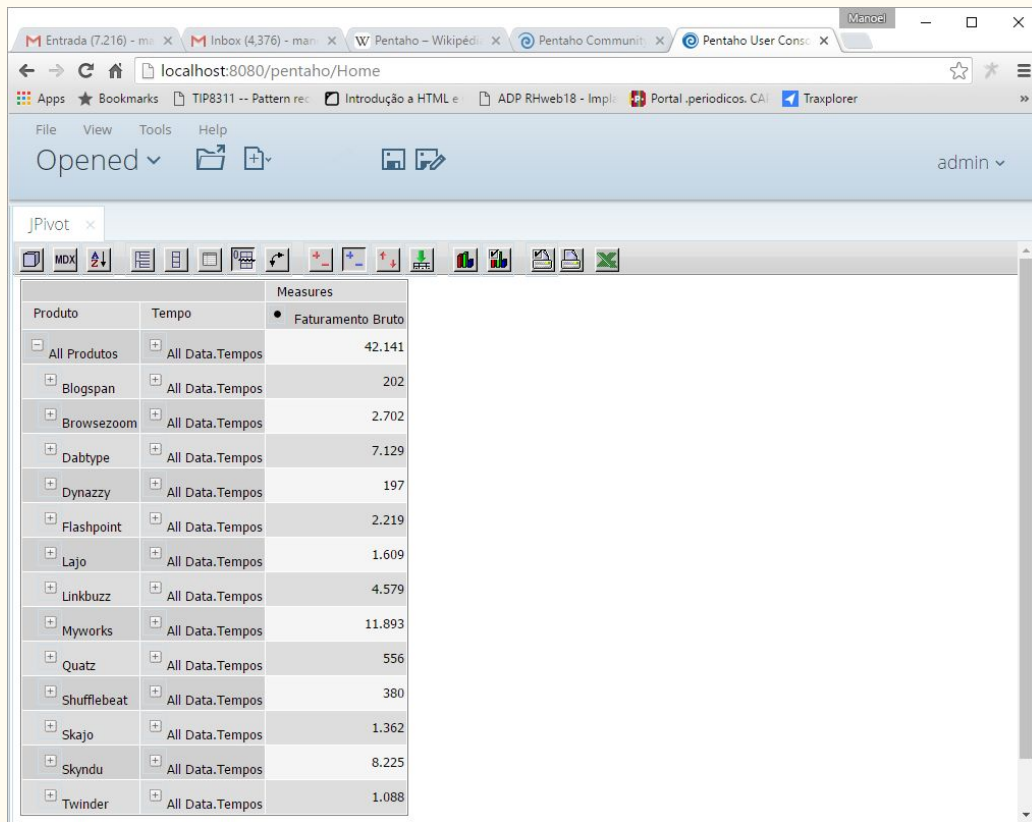
- Vendas
 - Table: fact_vendas
 - Produto
 - Produto
 - Categoria**
 - Item
 - Table: dim_item
 - Data
 - Faturamento Bruto
 - Desconto
 - Custo Direto
 - Custo Indireto
 - CM Lucro

Level for 'Produto' Hierarchy

Attribute	Value
name	Categoria
description	
table	
column	categoria
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	String
internalType	
uniqueMembers	<input type="checkbox"/>
levelType	Regular
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Database - DW (PostgreSQL)

Pentaho Business Analytics - JPivot



The screenshot displays the Pentaho JPivot web application running in a browser. The browser's address bar shows the URL `localhost:8080/pentaho/Home`. The application interface includes a top navigation bar with a 'File' menu, 'View' and 'Tools' buttons, and a 'Help' link. Below this is a 'Opened' section with folder icons and a user profile 'admin'. The main content area is titled 'JPivot' and features a toolbar with various icons for data manipulation. The data is presented in a table with three columns: 'Produto' (Product), 'Tempo' (Time), and 'Measures' (Measures). The 'Measures' column contains a single measure, 'Faturamento Bruto' (Gross Revenue). The table lists 15 products, each with a corresponding revenue value. The products are: All Produtos, Blogspan, Browsezoom, Dabtype, Dynazzy, Flashpoint, Lajo, Linkbuzz, Myworks, Quatz, Shufflebeat, Skajo, Skyndu, and Twinder.

Produto	Tempo	Measures
		• Faturamento Bruto
All Produtos	All Data.Tempos	42.141
Blogspan	All Data.Tempos	202
Browsezoom	All Data.Tempos	2.702
Dabtype	All Data.Tempos	7.129
Dynazzy	All Data.Tempos	197
Flashpoint	All Data.Tempos	2.219
Lajo	All Data.Tempos	1.609
Linkbuzz	All Data.Tempos	4.579
Myworks	All Data.Tempos	11.893
Quatz	All Data.Tempos	556
Shufflebeat	All Data.Tempos	380
Skajo	All Data.Tempos	1.362
Skyndu	All Data.Tempos	8.225
Twinder	All Data.Tempos	1.088

Fim