

# Modelagem de Dados no Pentaho BI

por Francisco Ernesto Teixeira - [fco.ernesto@gmail.com](mailto:fco.ernesto@gmail.com)

criado em 21/05/2013 06:26h

Este guia tem o objetivo de fazer uma simples modelagem de dados e publicá-lo no Pentaho BI. O mesmo foi escrito supondo que se esteja utilizando o Microsoft Windows. Entretanto deve-se lembrar que todos os aplicativos utilizados são multi-plataforma, com exceção do HeidiSQL que executa através do Wine (no Linux).

1. Servidor MySQL.....	1
2. MySQL Workbench.....	1
3. HeidiSQL.....	2
4. Suíte Pentaho BI Community Edition.....	2
4.1. Administração.....	2
4.1.1. Usuários e Regras.....	4
4.1.2. Conexões ao Banco de dados.....	5
4.2. Schema Workbench.....	5
4.2.1. Adicionando um Cubo.....	7
4.2.1.1. Adicionando uma Dimensão.....	8
4.2.1.1.1. Adicionando uma Hierarquia.....	9
4.2.1.1.1.1. Adicionando um Nível.....	10
4.2.1.2. Adicionando a dimensão Tempo.....	11
4.2.1.3. Adicionando um Indicador.....	11
4.2.2. Publicando o esquema mapeado.....	12
4.2.3. Atualizando o servidor com o esquema publicado.....	14

## 1. Servidor MySQL

Não vou explicar a sua instalação mas devo ressaltar que o mesmo deve estar instalado para que este guia seja concluído com sucesso. A suíte Pentaho BI funciona com a maioria dos bancos que esteja acessível por JDBC (MsSQL, PostgreSQL, Oracle, Firebird, entre outros).

## 2. MySQL Workbench

É necessário para modelar a solução estrela a ser utilizada.

Uma vez modelado a solução de BI, siga para o menu **Database**, submenu **Forward Enginner...**,

defina as configurações da conexão válidos para criar o banco e suas respectivas informações (dimensões, fatos, usuários, o que for necessário).

Clique em **Next**, clique novamente em **Next**, de novo em **Next** e finalmente em **Next**. O banco será criado na instância do servidor MySQL.

Para fechar a janela, clique em **Close**. E para encerrar o programa, siga para o menu **File**, submenu **Exit**.

### 3. HeidiSQL

Utilizado para alimentar o banco recém-criado com as informações as quais serão trabalhadas na análise do Pentaho BI.

Abrindo o programa, na janela **Gerenciar Sessões**, crie uma nova sessão com a instância do servidor MySQL com o usuário definido no MySQL Workbench.

Escolha o banco recém-criado e alimente suas dimensões e fato com as informações a serem trabalhadas.

## 4. Suíte Pentaho BI Community Edition

Uma suíte gratuita de código-livre para trabalhar com dataware house. A suíte é composta por inúmeras aplicações mas vamos somente dispor das ferramentas de administração, mapeamento do esquema e console do aplicativo.

### 4.1. Administração

Para utilizar a modelagem de dados que criamos no Pentaho BI é necessário criar os usuários que irão ter acesso e a conexão da instância do servidor MySQL com o mesmo.

Deve-se, antes de tudo, iniciar seu Console. Sem ele nada do que será feito funcionará.

Clique no botão Iniciar, vá para o menu **Todos os programas**, submenu **Pentaho Community Edition**, submenu **Server Management** e clique em **Start Business Intelligence**.

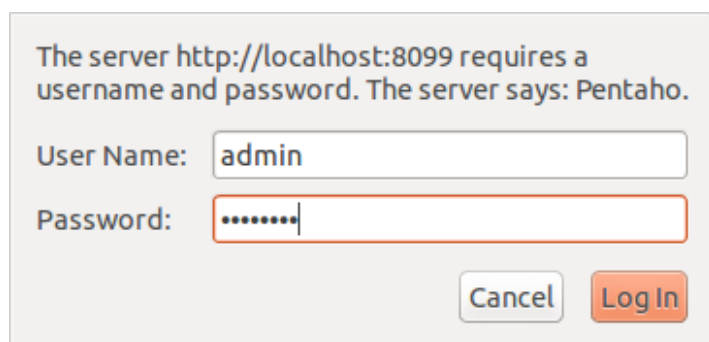
Quando na tela de terminal aparecer a mensagem **Server startup . . .** o servidor já estará em plena atividade.

Precisamos iniciar o servidor da administração. Clique no botão Iniciar, vá para o menu **Todos os programas**, submenu **Pentaho Community Edition**, submenu **Server Management** e clique em **Start Administration Console**.

Quando na tela de terminal aparecer a mensagem **Console is now started.** o servidor já estará em plena atividade.

Para abrir a janela da administração, clique no botão Iniciar, vá para o menu **Todos os programas**, submenu **Pentaho Community Edition** e clique em **Administration Console Login**.

Quando abrir a janela da administração, irá ser solicitado o usuário e senha, que são **admin** e **password** respectivamente. A administração irá ser inicializada.



The server http://localhost:8099 requires a username and password. The server says: Pentaho.

User Name:

Password:

No lado esquerdo, existem as opções **Home** e **Administration**. Clique em **Administration**.



Aparecerá as abas **Users & Roles**, **Database Connections**, **Services** e **Scheduler**.

Iremos nos limitar somente à aba **User & Roles** e **Database Connections**.

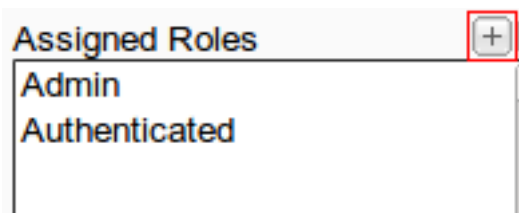
#### 4.1.1. Usuários e Regras

Na aba **User & Roles** clique em **Users**. Em **Users** existe próximo ao nome em cima da listagem um símbolo de mais. Com ele você pode adicionar um usuário.



Clique no símbolo de mais e aparecerá a janela **Add User**. Forneça as informações necessárias e clique em **OK**.

Clique em **Roles**. Clique na regra **Admin**. Próximo ao rótulo **Assigned Users** existe um símbolo de mais. Com ele você pode adicionar o usuário que você criou à regra selecionada.



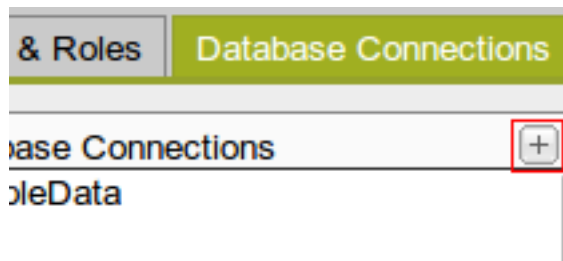
Clique no símbolo de mais e aparecerá a janela **Assign Users**. Na listagem **Available** constará o usuário que você criou. Selecione o usuário que você criou e no botão **> (seta para direita)**. Clique depois no botão **OK**.

Acabamos de criar o usuário que será usado no Pentaho BI.

## 4.1.2. Conexões ao Banco de dados

Clique na aba **Database Connections**.

Próximo ao rótulo **Database Connections** existe um símbolo de mais. Com ele você pode adicionar conexões ao banco de dados que será utilizado pelo Pentaho BI para acessar o banco que criamos.



Clique no símbolo de mais e aparecerá a janela **Add Database Connection**. Preencha os campos da seguinte forma:

- Name: <<nome do banco de dados que será conectado>>
- Driver Class: **org.gjt.mm.mysql.Driver**
- User Name: <<usuário com acesso ao banco de dados>>
- Password: <<senha do usuário com acesso ao banco de dados>>
- URL: **jdbc:mysql://<<endereço do servidor>>:<<porta da instância MySQL>>/<<banco de dados>>**

Clique no botão **Test** para verificar se todas as informações foram preenchidas corretas.

Finalizamos as configurações na administração do Pentaho BI. A administração será utilizada novamente após ser concluído o mapeamento, onde serão recarregados o cachê do mapeamento.

## 4.2. Schema Workbench

Criado o usuário e estabelecida a conexão do Pentaho BI, é preciso mapear as informações do banco que foi criado dizendo o comportamento de suas tabelas. Isso é feito com o Schema Workbench.

Clique no botão Iniciar, vá para o menu **Todos os programas**, submenu **Pentaho Community Edition**, submenu **Design Tools** e clique em **Schema Workbench**.

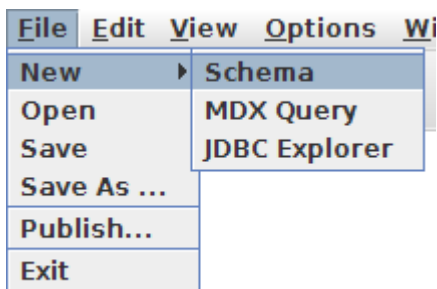
No Schema Workbench, siga para o menu **Options** e **Connection**. Aparecerá a janela **Database Connection**. Forneça as seguintes informações:

- Connection Name: <<nome que você deseja dar para a conexão>>
- Connection Type: **MySQL**
- Access: **Native (JDBC)**
- Settings > Host Name: <<endereço do servidor>>
- Settings > Database Name: <<banco de dados>>
- Settings > Port Number: <<porta da instância MySQL>>
- Settings > User Name: <<usuário com acesso ao banco de dados>>
- Settings > Password: <<senha do usuário com acesso ao banco de dados>>

Clique no botão **Test** para validar as informações fornecidas e finalmente clique no botão **OK**.

O programa aparentemente vai travar, mas em verdade ele está propagando a conexão fornecida por ele.

Uma vez que ele volte a responder, clique no menu **File**, submenu **New** e escolha **Schema**.

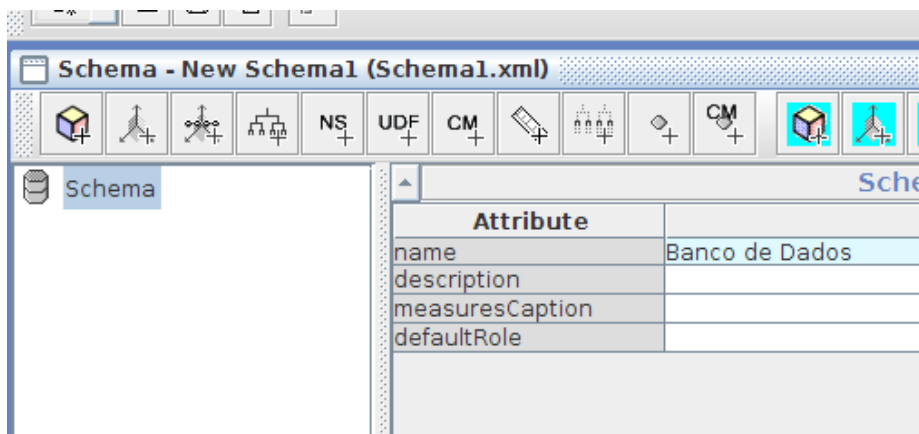


Para evitar a perda de informações, deve-se manter salvando o mapeamento que está sendo feito. Clique no menu **File** e escolha "**Save As ...**".

Forneça o local e o nome do arquivo. Uma sugestão é o nome seja o nome do banco que você criou.

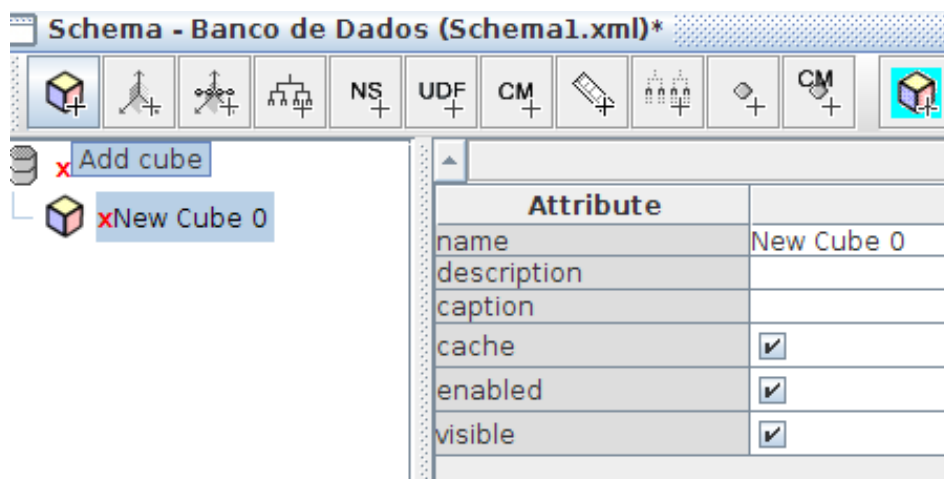
Já está pronto para o mapeamento.

Clique em **Schema** e no atributo **name** coloque o nome do banco de dados que você criou.



### 4.2.1. Adicionando um Cubo

Depois, com o esquema selecionado clique no botão **Add Cube**.

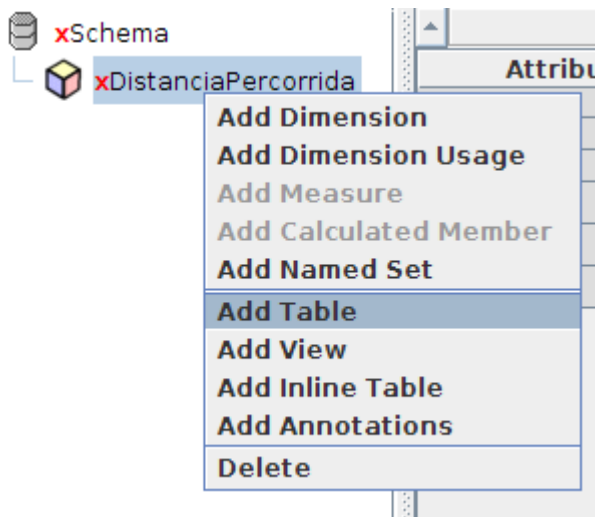


No atributo **name** do novo cubo, coloque o nome do cubo que será utilizado no Pentaho BI. Podem haver inúmeros cubos.

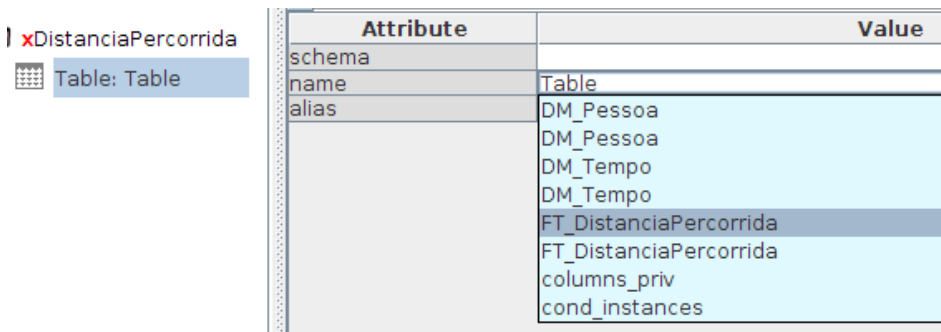
Nesse caso particularmente o nome do cubo é *DistanciaPercorrida*. O importante é observar as mensagens em vermelho que são exibidas na parte inferior da janela. A mensagem deste momento é **Fact name must be set** que quer dizer que precisamos especificar uma tabela da base de dados que será a fato deste cubo.

Sempre que aparecer uma mensagem é porque falta alguma informação e torna o cubo defeituoso.

Clique com o botão direito do mouse em cima do cubo e selecione **Add Table**.

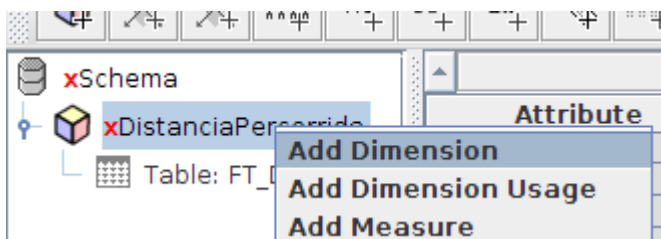


Adicionada a tabela, selecione-a e no seu atributo **name** escolha o nome da tabela fato.



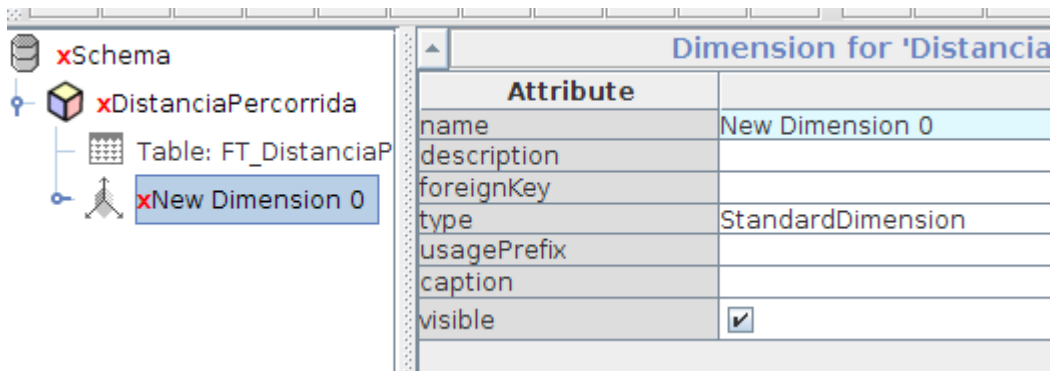
#### 4.2.1.1. Adicionando uma Dimensão

Selecione o cubo. Clicando novamente com o botão direito do mouse em cima do cubo, deve-se escolher **Add Dimension**.



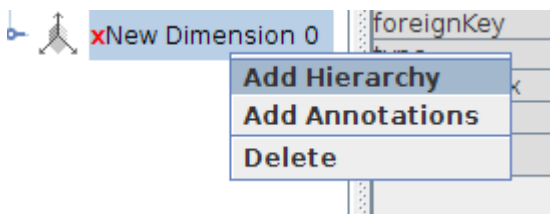
Será adicionado o elemento **New Dimension 0**, que é o mapeamento para as dimensões do modelo. Dentro dele devem existir outros elementos.



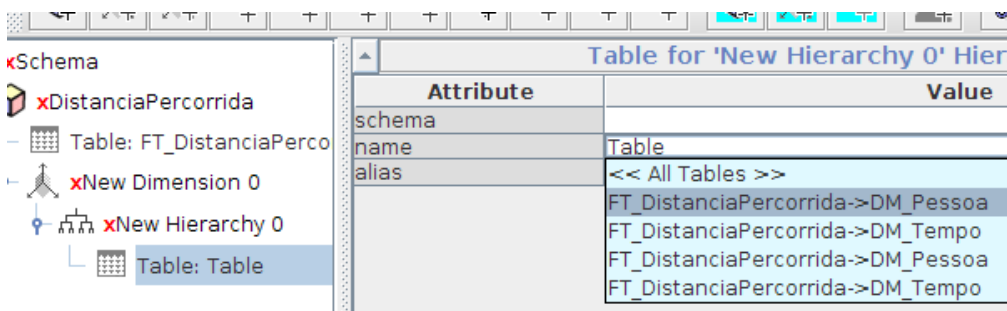


#### 4.2.1.1.1. Adicionando uma Hierarquia

Adicione uma hierarquia.

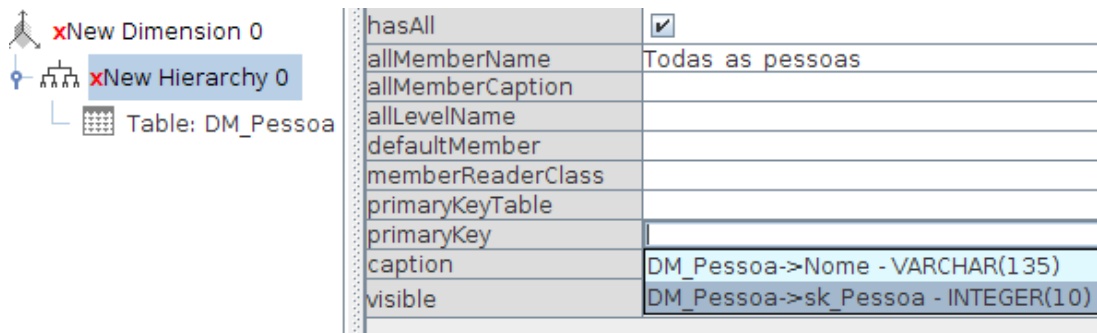


E na hierarquia, adicione uma tabela. Que será a tabela que é uma dimensão no banco de dados. Uma vez adicionada a tabela, no atributo **name** selecione-a tal como foi feito na fato. Lembre-se de deixar o atributo **schema** vazio. Ele não se aplica ao MySQL e causará problemas no próximo passo.

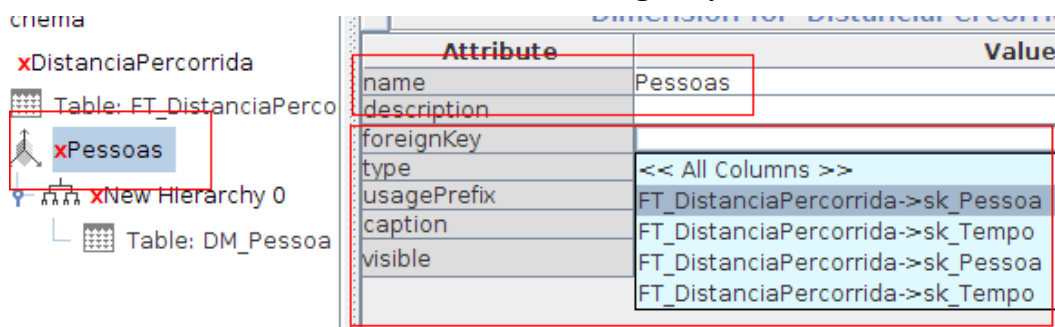


Selecione novamente a hierarquia. No atributo **allMemberName** forneça o nome, neste caso, **Todas\_as\_Pessoas**. Este é o nome que aparecerá nos relatórios, análise, etc.

É preciso também, nesta hierarquia, selecionar a chave primária da tabela que foi adicionada. O mesmo é feito através do atributo **primaryKey**.

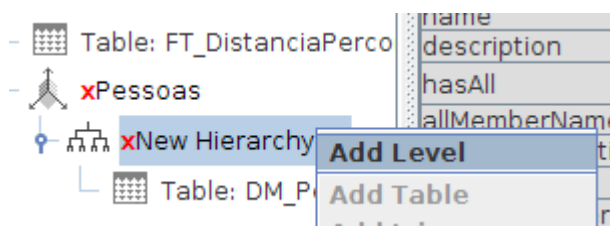


Selecione novamente a dimensão, preencha o atributo **name** e escolha a chave do relacionamento entre as tabelas fato e dimensão no atributo **foreignKey**.

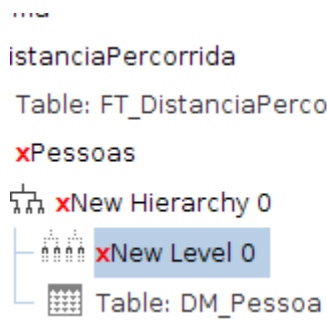


#### 4.2.1.1.1. Adicionando um Nível

As tabelas fato e dimensão foram devidamente mapeadas. Mas na hierarquia não foi selecionado o(s) campo(s) que serão exibidos. Para isto, selecionado a hierarquia, clique com o botão direito do mouse e escolha **Add Level**.



Adicionado o nível, deve-se em seu atributo **name** colocar o nome que será exibido. No atributo **column** deve-se escolher o campo da tabela que será exibido.



istanciaPercorrida

Table: FT\_DistanciaPerco

xPessoas

xNew Hierarchy 0

xNew Level 0

Table: DM\_Pessoa

Attribute	Value
name	New Level 0
description	
table	
column	
nameColumn	DM_Pessoa->Nome - VARCHAR(135)
parentColumn	DM_Pessoa->sk_Pessoa - INTEGER(10)
nullParentValue	
ordinalColumn	

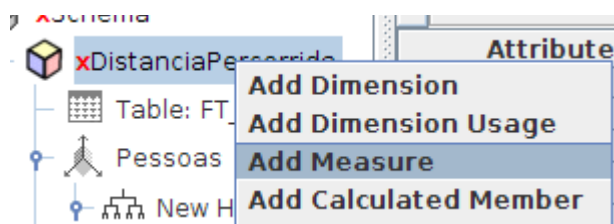
#### 4.2.1.2. Adicionando a dimensão Tempo

Para adicionar a dimensão Tempo, deve-se seguir os mesmos passos de 4.2.1.1. *Adicionando uma Dimensão*. A única diferença é que o mesmo possuirá mais níveis. Exemplo: nível Ano, nível Trimestre, nível Mês, entre outros.

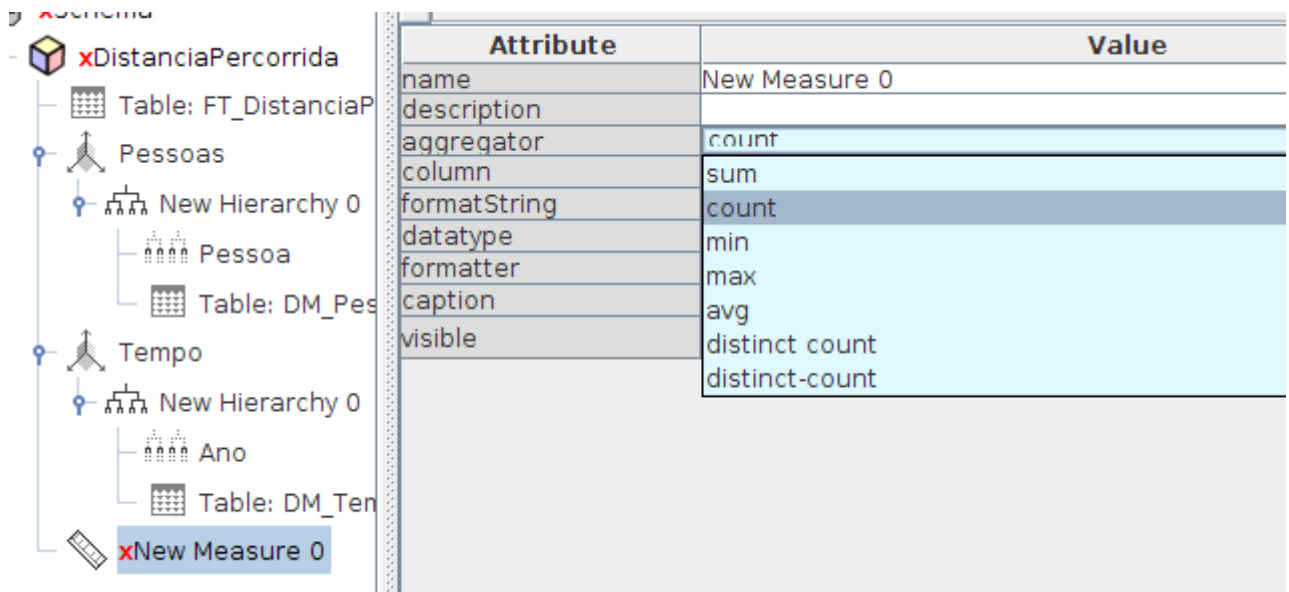
#### 4.2.1.3. Adicionando um Indicador

Todo cubo necessita de um indicador, ou indicadores, que são a média, total, soma dos valores do fato, entre outros indicadores.

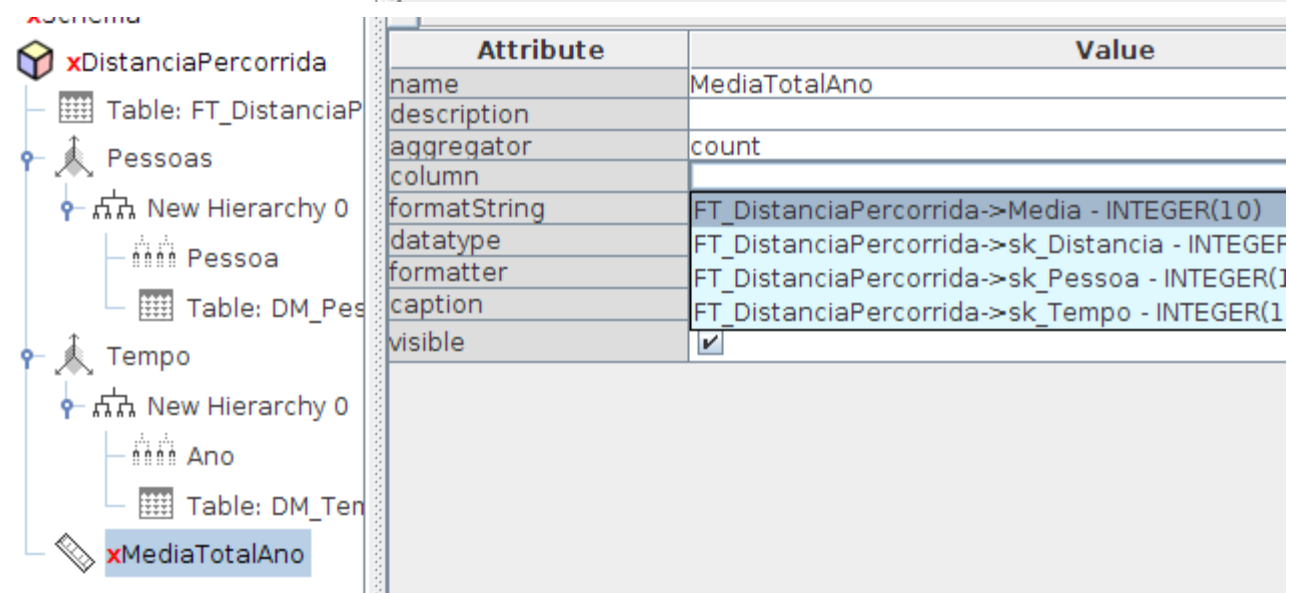
Para adiciona um indicador, selecione o cubo, clique com o botão direito do mouse e escolha **Add Measure**.



Quando for adicionado o indicador, ele possui entre seus principais atributos o **name**, **column** e **aggregator**. O **aggregator** escolhido foi a **soma (sum)**.



Attribute	Value
name	New Measure 0
description	
aggregator	count
column	sum
formatString	count
datatype	min
formatter	max
caption	avg
visible	distinct count
	distinct-count



Attribute	Value
name	MediaTotalAno
description	
aggregator	count
column	
formatString	FT_DistanciaPercorrida->Media - INTEGER(10)
datatype	FT_DistanciaPercorrida->sk_Distancia - INTEGER(10)
formatter	FT_DistanciaPercorrida->sk_Pessoa - INTEGER(10)
caption	FT_DistanciaPercorrida->sk_Tempo - INTEGER(10)
visible	<input checked="" type="checkbox"/>

#### 4.2.2. Publicando o esquema mapeado

Não esqueça de ficar salvando o esquema enquanto estiver mapeando.

Uma vez mapeado, podemos publicar no Pentaho BI. Para isso devemos ir ao menu **File**, submenu **Publish...**.

Irá aparecer a janela **Repository Login**. Escolha o servidor, forneça a senha para publicar – se uma

instalação recente: **password** – e forneça o usuário e senha que você criou na administração e que tenha permissão para poder publicar. No final clique no botão **OK**.

The dialog box contains two main sections. The 'Server' section has a 'URL:' label followed by a text box containing 'http://localhost:8080/pentaho/' and a dropdown arrow, and a 'Publish Password:' label followed by an empty text box. The 'Pentaho Credentials' section has a 'User:' label followed by an empty text box and a 'Password:' label followed by an empty text box. At the bottom, there is a checked checkbox labeled 'Remember these Settings' and two buttons: 'OK' and 'Cancel'.

**Server**  
URL:  
  
Publish Password:

**Pentaho Credentials**  
User:  
  
Password:

☒ Remember these Settings

**OK** **Cancel**

Após clicar no botão **OK**, aparecerá uma janela notificando que está tentando acessar o servidor com as permissões fornecidas. A janela **Publish Schema** aparecerá.

The dialog box displays schema information. It has labels for 'Schema Name:' (Banco de Dados), 'Schema File:' (Schema1.xml), and 'Location:' followed by a text box containing a file path and icons. Below this is a table with three columns: 'Name', 'Type', and 'Date Modified'. The table lists several folders. At the bottom, there is a 'Publish Settings' section with a 'Pentaho or JNDI Data Source:' label and a text box containing 'SampleData', and a checkbox labeled 'Register XMLA Data Source'. Two buttons, 'Publish' and 'Cancel', are at the bottom right.

**Schema Name:**  
*Banco de Dados*  
**Schema File:**  
*Schema1.xml*  
**Location:**

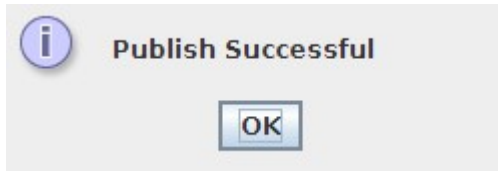
Name	Type	Date Modified
admin	File Folder	03/19/2013 04:16 PM
bi-developers	File Folder	03/19/2013 04:16 PM
cdb	File Folder	12/31/1969 09:00 PM
cde	File Folder	12/31/1969 09:00 PM
cdv	File Folder	12/31/1969 09:00 PM
plugin-samples	File Folder	03/19/2013 04:16 PM
steel-wheels	File Folder	03/19/2013 04:16 PM
teste	File Folder	03/19/2013 04:18 PM

**Publish Settings**  
**Pentaho or JNDI Data Source:**  
  
☐ Register XMLA Data Source

**Publish** **Cancel**

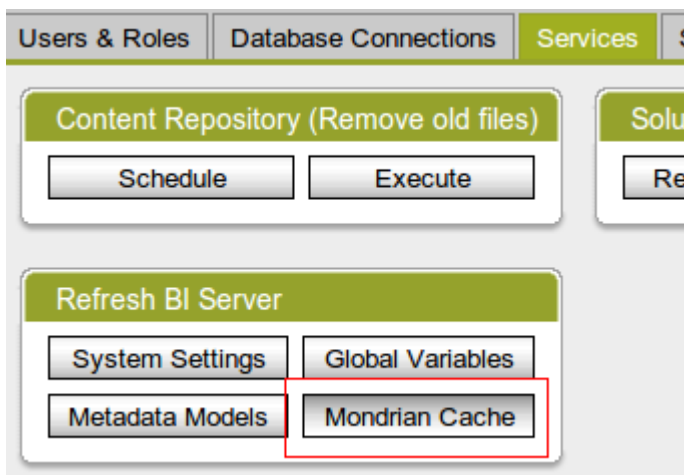
Escolha o local para publicar o esquema, nesse caso **teste**, e no **Pentaho or JNDI Data Source** forneça o nome da conexão que foi criado na administração.

Ao final, clique no botão **Publish**. Será apresentado com uma mensagem de sucesso.



### 4.2.3. Atualizando o servidor com o esquema publicado

Concluído o mapeamento do esquema, é necessário atualizar o Pentaho BI. Para isso, acesse novamente a administração e vá na aba **Services**. Clique no botão **Mondrian Cache**.



Com isso é certificado que o esquema foi devidamente publicado e atualizado.

## 4.2. Console do Pentaho BI: Onde a mágica acontece

Já foi aprendido a modelar, popular, mapear e publicar os dados no Pentaho BI. É preciso agora fazer as análises em si.

Para abrir a janela do console, clique no botão Iniciar, vá para o menu **Todos os programas**, submenu **Pentaho Community Edition** e clique em **Pentaho User Console Login**. Forneça o

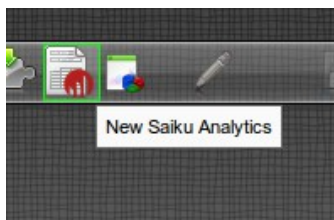
usuário e senha criados na administração e clique no botão **Login**.



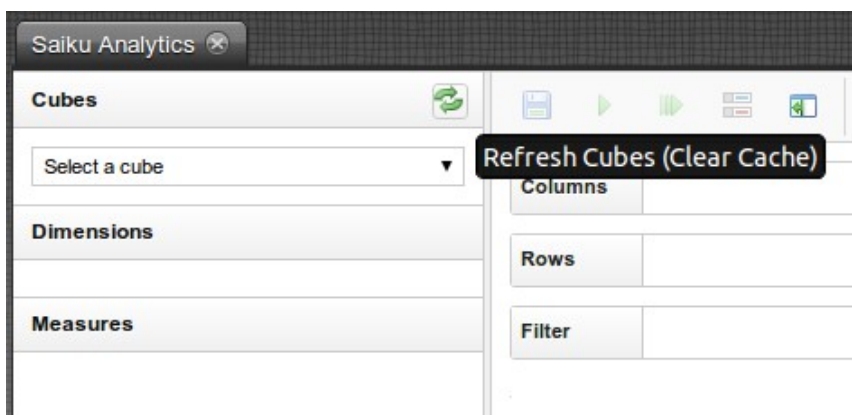
[www.pentaho.com](http://www.pentaho.com) | [Contact Us](#)



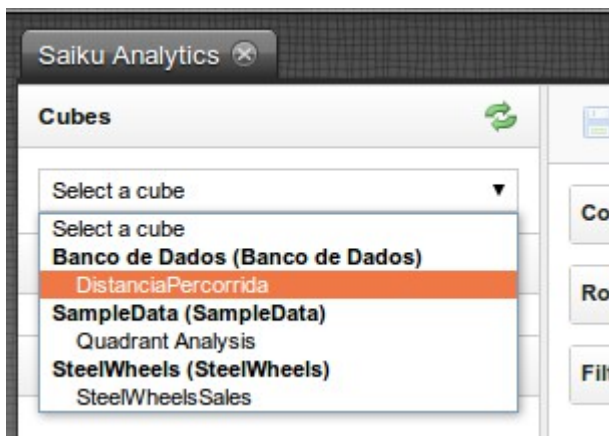
Na tela do Pentaho BI, na barra superior, clique no botão **Saiku Analytics**.



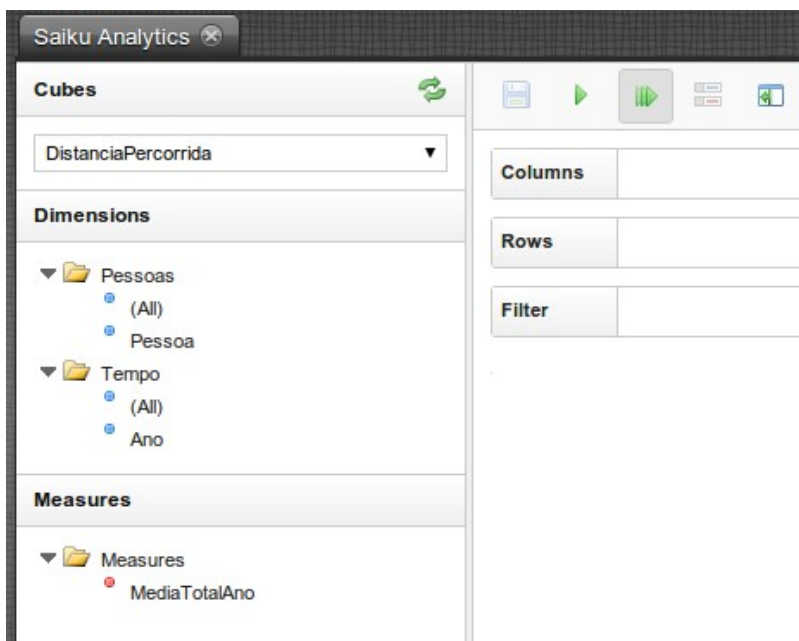
No formulário do **Saiku Analytics**, clique no botão de recarregar para atualizar a lista dos cubos disponíveis.



Selecione um dos cubos que foi publicado pelo mapeamento.



Será apresentado com as dimensões e indicadores do cubo.

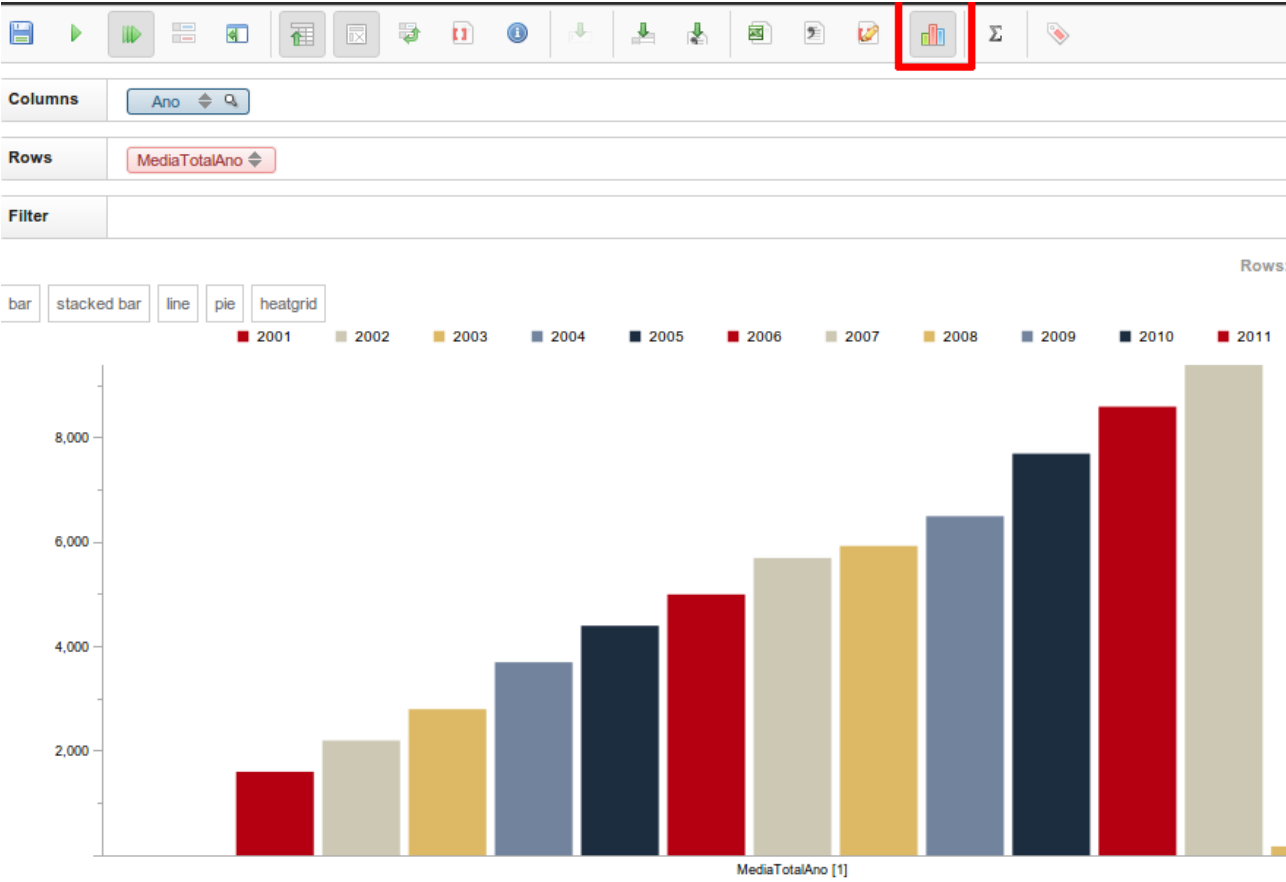


Clique, segure, arraste e solte as dimensões e indicadores para serem analisados.

<div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div> </div>											
Columns	Ano										
Rows	MediaTotalAno										
Filter											
MeasuresLevel	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
MediaTotalAno	1,600	2,200	2,800	3,700	4,400	5,000	5,700	5,930	6,500	7,700	8,600



Clicando nos botões que acompanham a ferramenta, os resultados podem ser apresentados de diferentes formas e salvos em diferentes formatos.



Divirta-se!