

Desenvolvendo projeto de BI com Pentaho

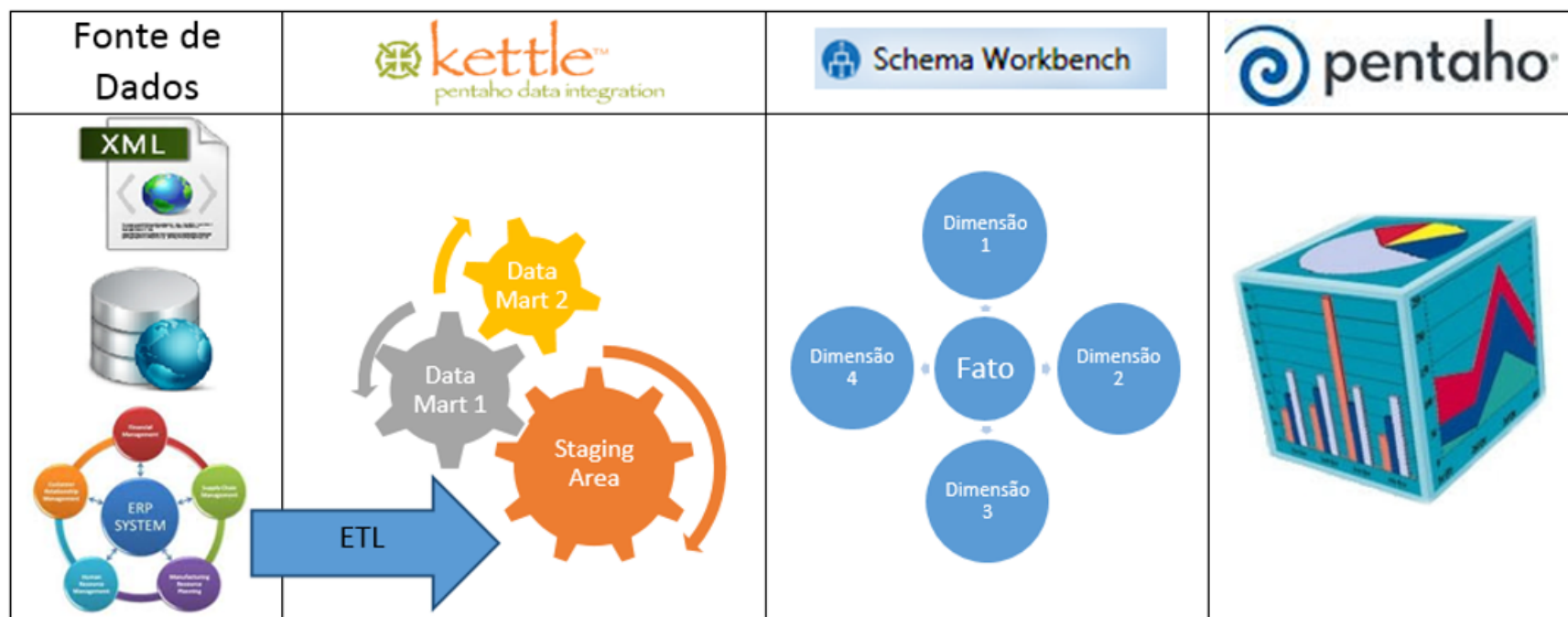
Por Vinícius

Oliveira

Neste artigo veremos o funcionamento geral do Pentaho no desenvolvimento de um projeto de Business Intelligence. A figura abaixo mostra na prática como funciona o processo de construção do *Data Warehouse* utilizando o Pentaho.



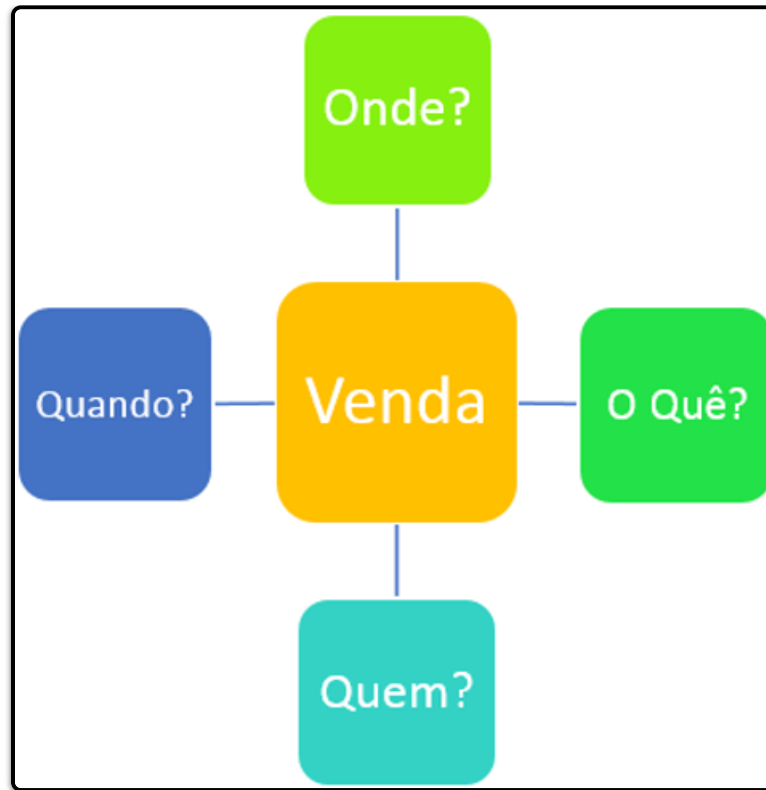
Clique aqui e participe



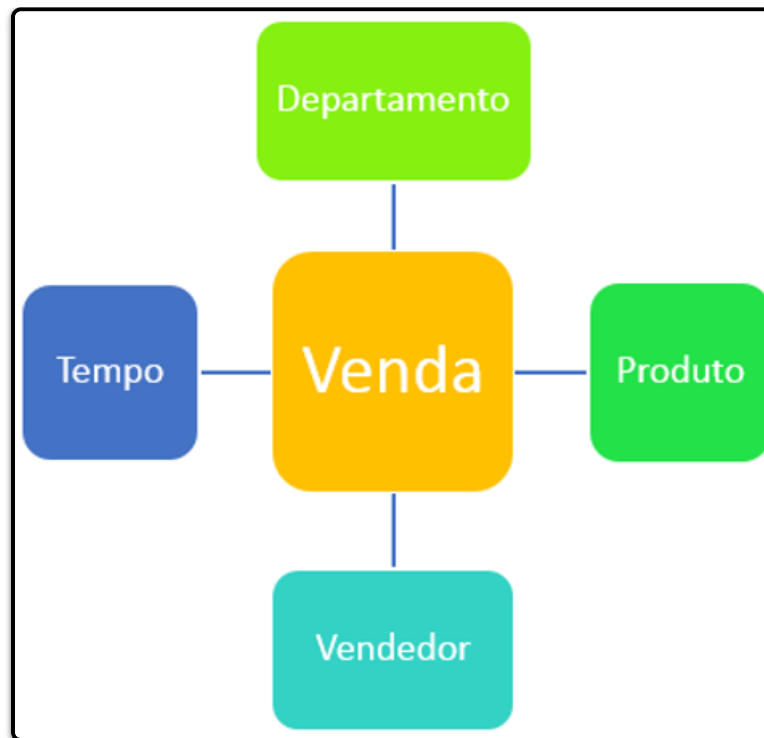
Construindo um *Data Mart* de Vendas

O Data Mart pode ser definido como um subconjunto de um Data Warehouse, normalmente são dados referente a um assunto especial (ex: Vendas, Estoque, Recursos Humanos).

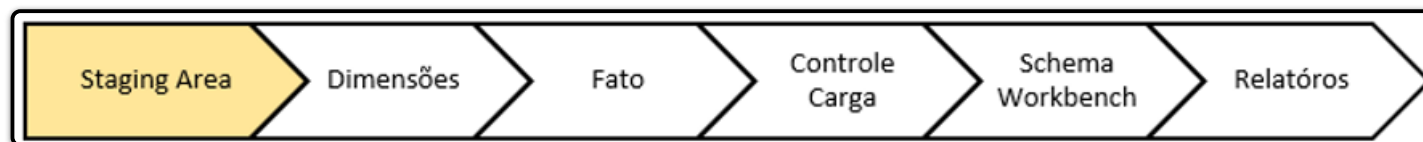
Após definir os requisitos necessários para criação do *Data Mart*, temos que criar o nosso modelo multidimensional. No livro "Tecnologia e Projeto de *Data Warehouse*" (Felipe Nery R. Machado) o autor explica um modelo muito prático de como criar um modelo multidimensional. Foi dividido as dimensões em quatro tipos: Onde, O quê, Quando e Quem. Com esses tipos conseguimos desenvolver o nosso modelo multidimensional mais facilmente.



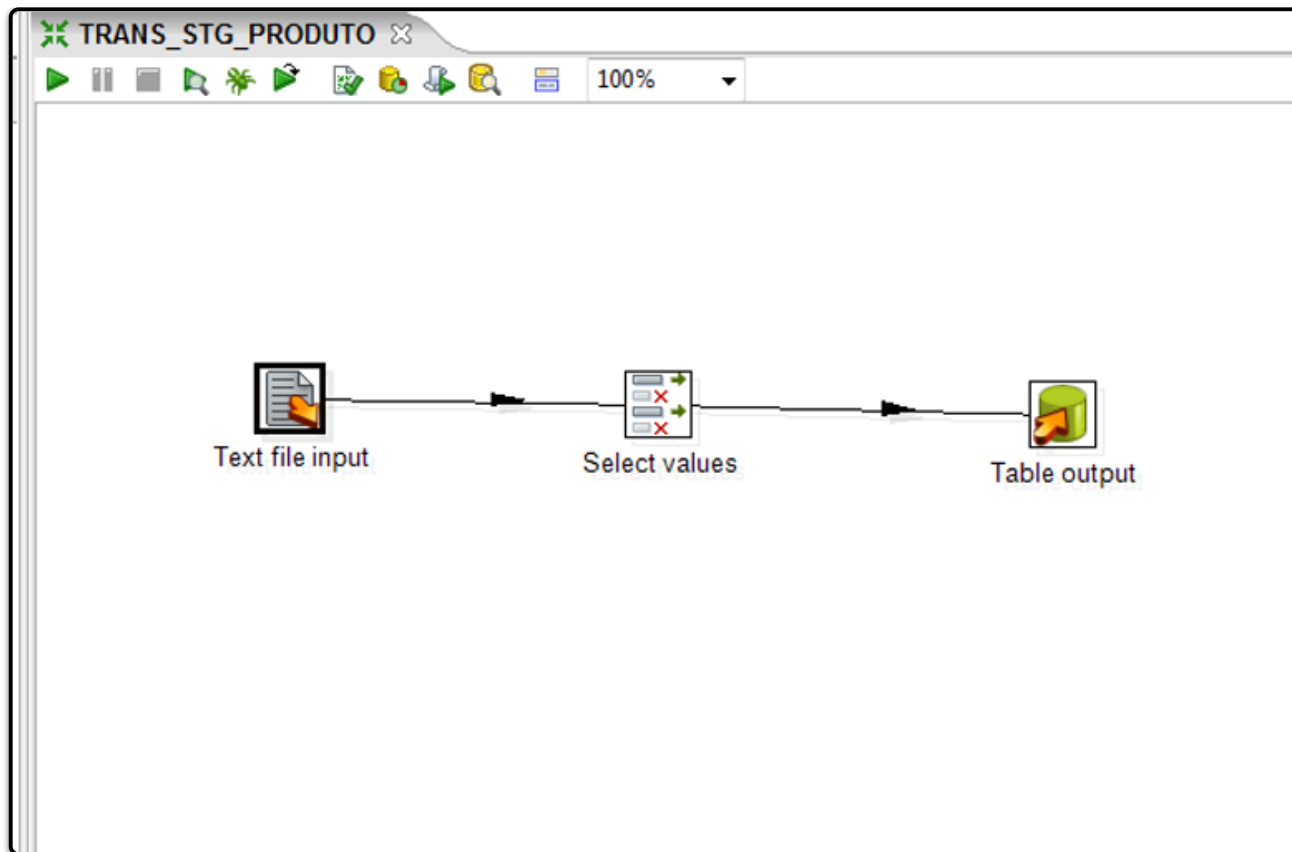
- Dimensão Onde: Onde foi realizada a venda? (Departamento)
- Dimensão O Quê: O que foi vendido? (Produto)
- Dimensão Quem: Quem realizou a venda? (Vendedor)
- Dimensão Quando: Quando foi realizado a venda? (Tempo)



Carregando *Staging Area*

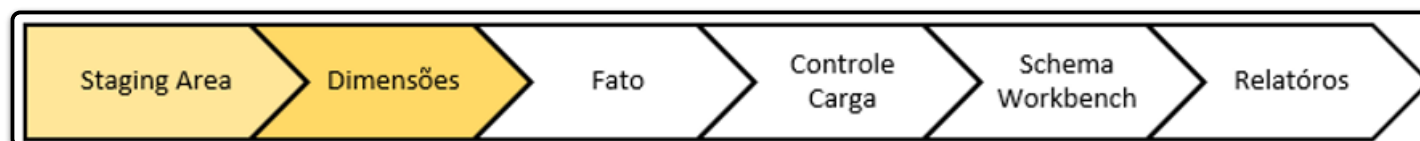


A área de *staging* é local de armazenamento dos dados intermediários, que fica entre a origem dos dados original e o repositório de dados centralizado. Todo tratamento, reconciliação, transformação e relacionamentos necessários acontecem nessa área. Essa implementação é muito bem definida e progressiva. Ter uma área de *staging* é uma estratégia comum e simples para obter informações de um sistema transacional sem a necessidade de realizar transformações complexas.



No exemplo acima, a *staging* é carregada a partir de um arquivo texto, utilizando o *step* de "*Text file input*" (Origem), "*Select values*" (Mapeamento) e "*Table output*" (Destino).

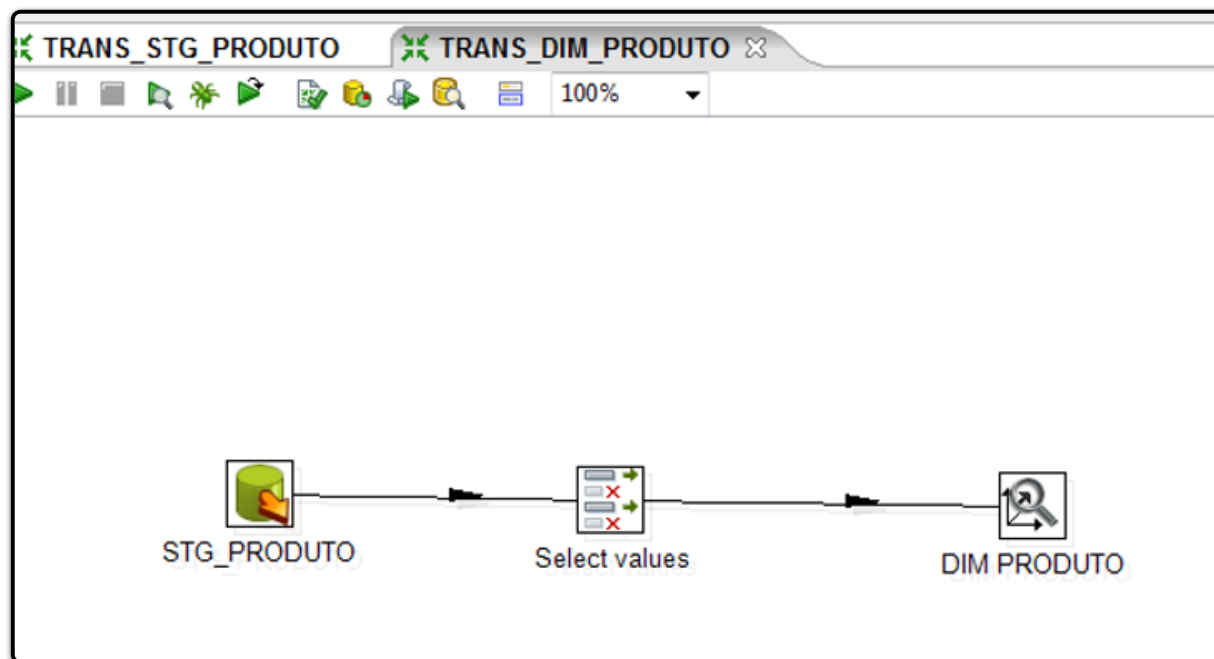
Carregando Dimensão



O Pentaho Data Integration, possui um *step* chamado "*Dimension Lookup/update*". Neste componente podemos definir os tipos de dimensões utilizadas com a implementação do SCD (*Slowly Changing Dimension*), mudando as dimensões para os tipos SCD 1, SCD 2, SCD 3 ou SCD híbrido (SCD 6).

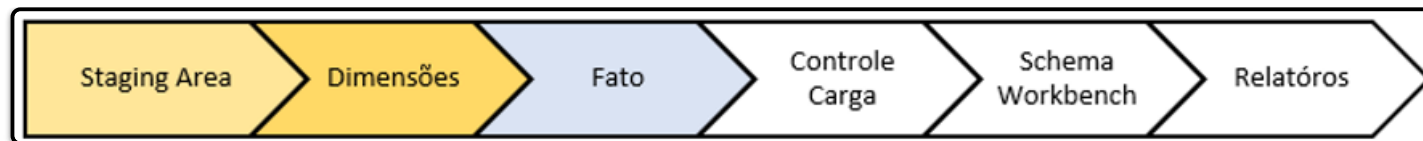
Você poderá entender melhor o SCD, neste artigo: **O que significa e qual a importância do SCD no Data Warehouse.**

Após entender um pouco mais sobre SCD vamos ver como fica nosso mapeamento utilizando o PDI:



Utilizamos a *staging* criada para realizar a carga da dimensão produto.

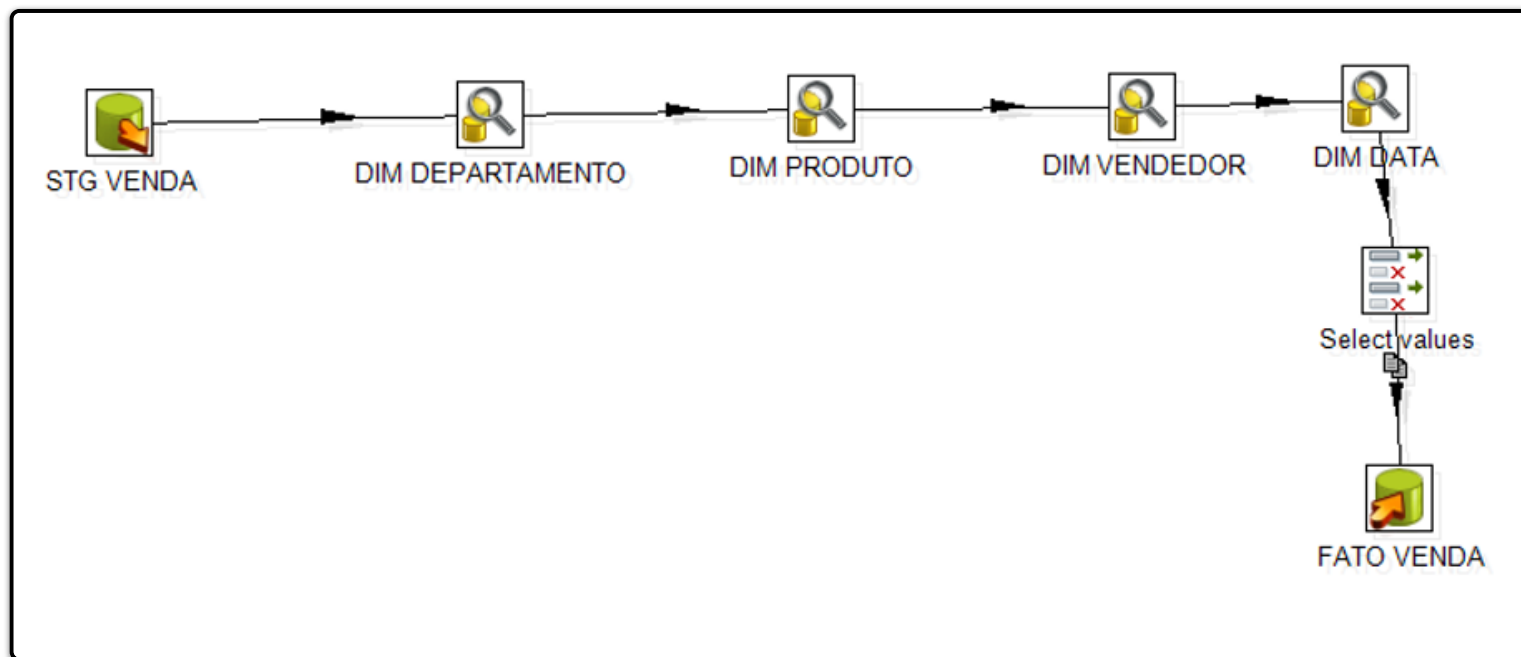
Carregando a Fato



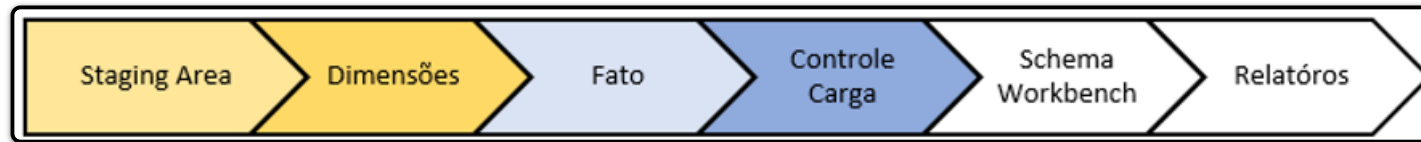
A fato venda, contém as informações de vendas de produtos de uma determinada empresa. Essa fato irá reunir as chaves artificiais (no inglês *surrogate keys*), métricas (valor da venda) para que possamos construir nossas análises.

Para obter as chaves artificiais iremos utilizar o *step "Database Lookup"*, onde a partir de uma chave, por exemplo código do vendedor, juntamente com a data da venda, irá retornar a *surrogate key* (SK) exata do registro.

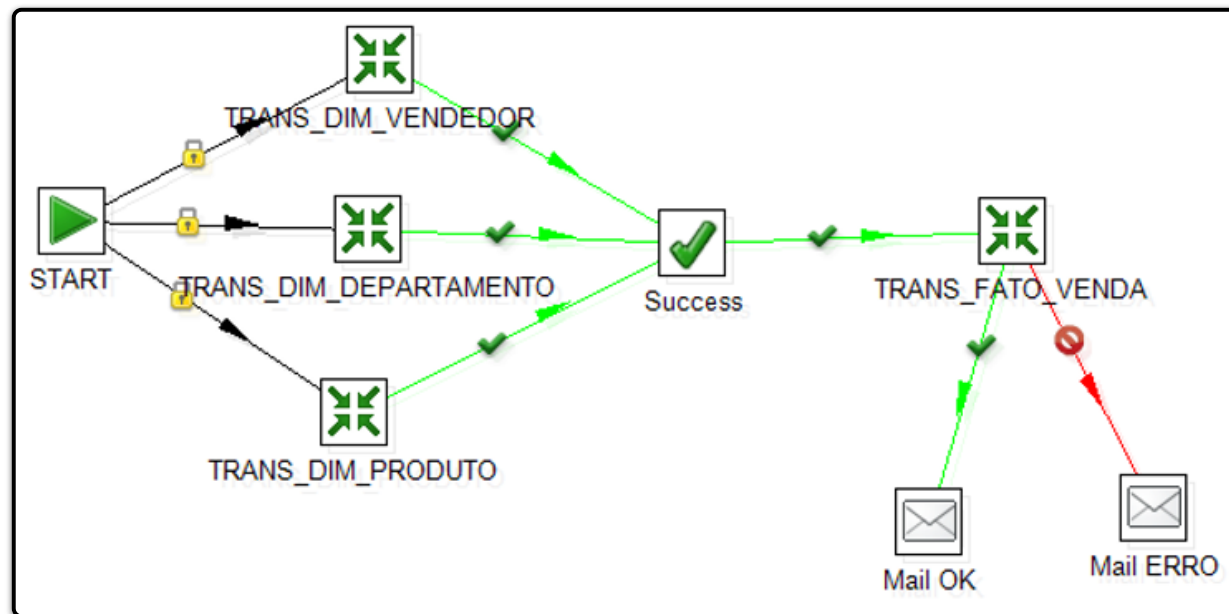
Abaixo é possível visualizar o mapeamento de carga da fato.



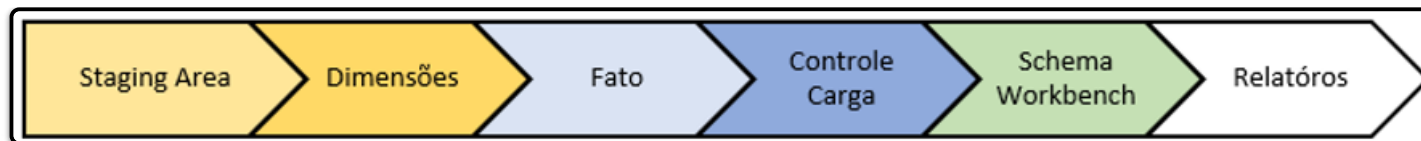
Controle Carga



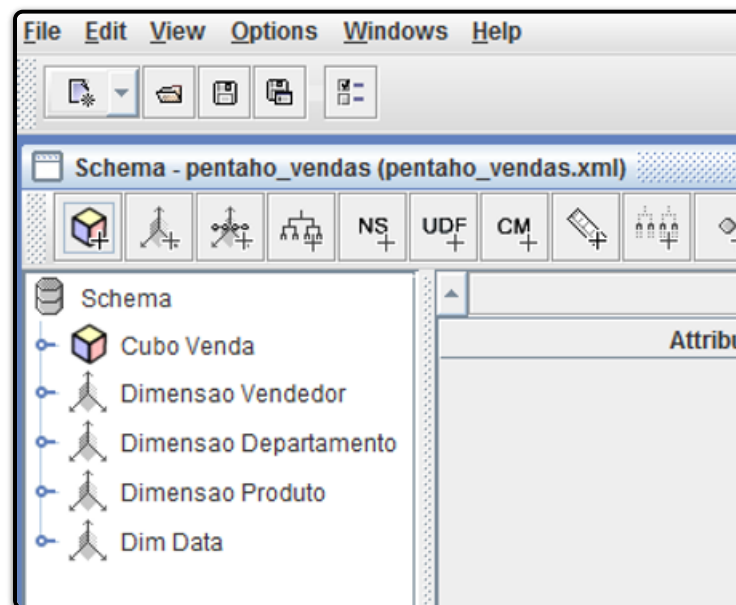
A etapa de controle carga, concentra a inteligência da carga em um único elemento. No nosso exemplo iremos realizar o controle de carga a partir de um *job* utilizando o PDI.



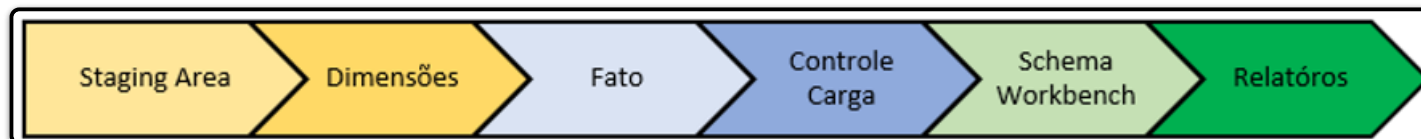
Configurando Schema Workbench



No *Schema Workbench* iremos configurar nossas dimensões e nosso cubo com as métricas. Essa ferramenta apresenta diversos componentes que permite configurar o cubo visualmente.

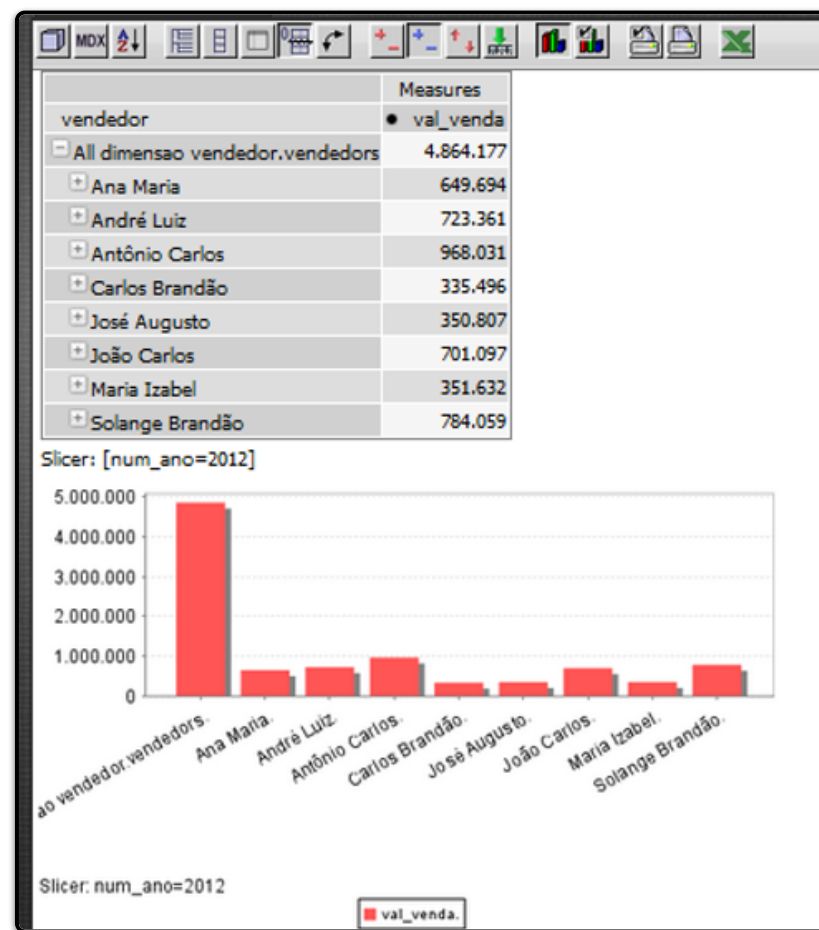


Relatório no *Pentaho User Console*



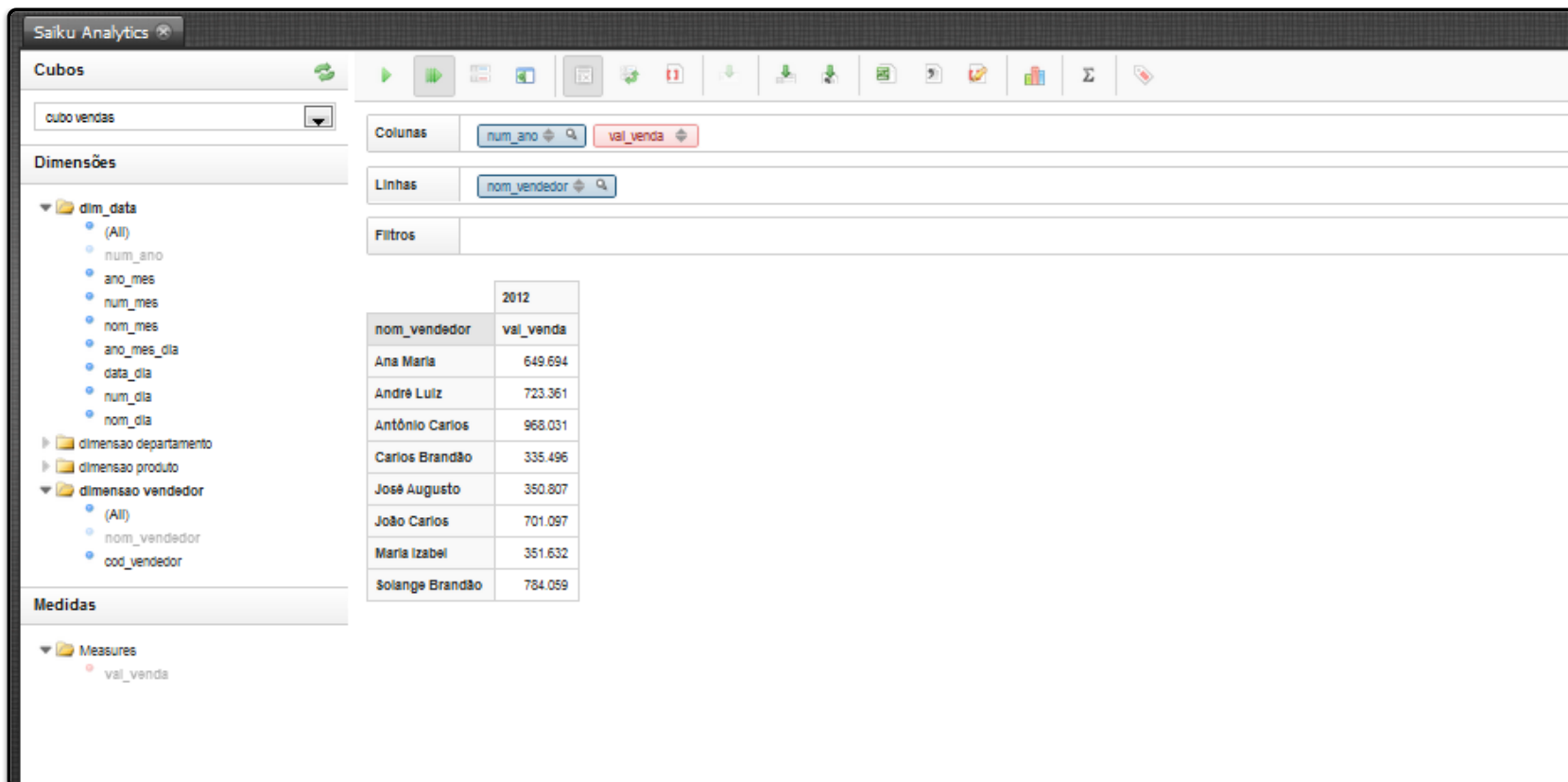
No PUC (*Pentaho User Console*) podemos realizar relatórios utilizando diversas ferramentas, como o *Saiku*, *Report Designer*, *Ctools* e *JPivot*.

JPivot é uma biblioteca que funciona a partir de tags JSP que permite ao usuário executar navegações OLAP típicas como *slice and dice*, *drill-down* e *roll-up*. Ele usa Modrian como seu servidor OLAP.



Saiku é uma suíte de código aberto que oferece uma poderosa ferramenta OLAP que permanece facilmente incorporável, extensível e configurável.

Com o Saiku é possível se conecta a sistemas existentes OLAP, a exemplo do Mondrian (Pentaho).



O **Pentaho Report Designer (PRD)** é uma aplicação *desktop* que proporciona um ambiente de *design* visual para criar definições de relatório. As definições de relatório podem ser executadas e salvo localmente através do PRD ou publicado em um servidor Pentaho BI para permitir que várias pessoas possam acessar e agendar a execução do relatório. O PRD é voltado para analistas e que estão familiarizados com os conceitos e as fontes de dados utilizadas.

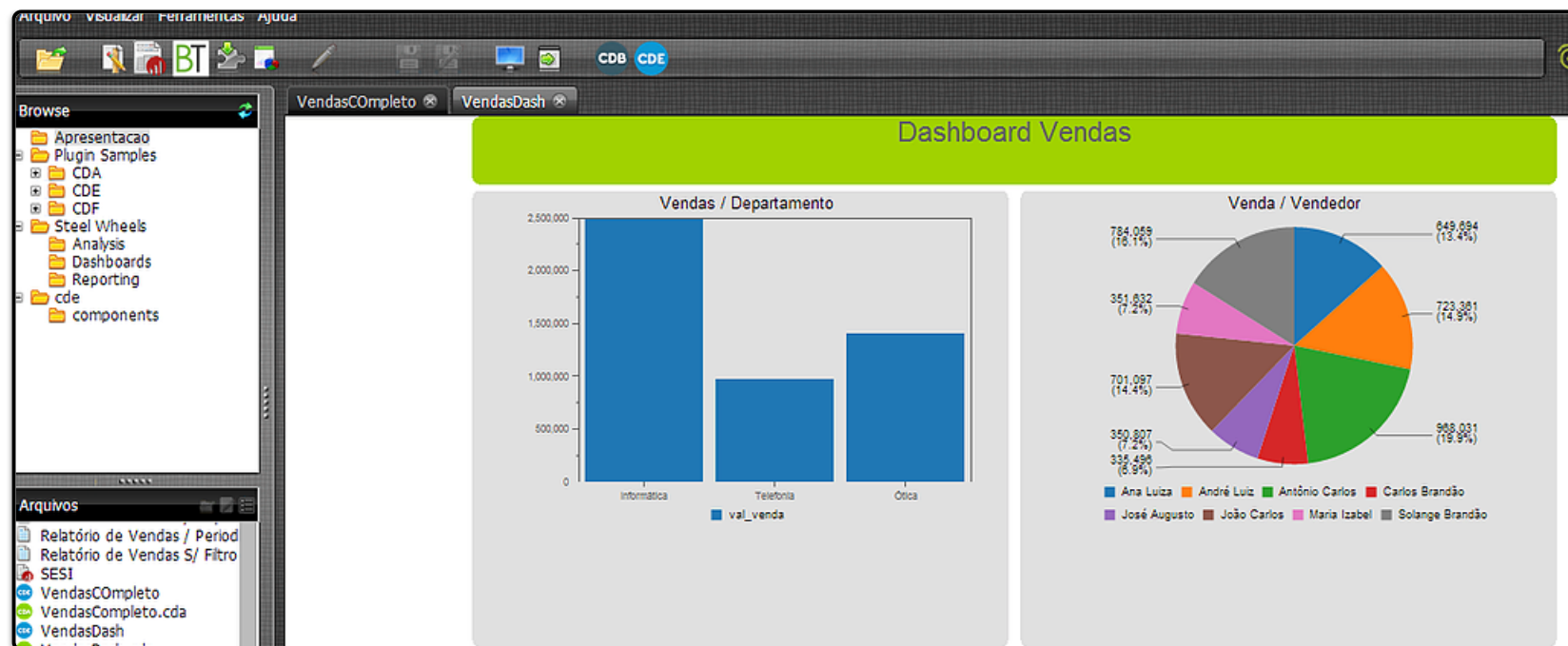
Além das análises o Pentaho também pode gerar relatórios para visualização mais detalhada dos dados. A ferramenta padrão para criação de relatórios é o Pentaho Report Designer (PRD).

O PRD é uma ferramenta com interface intuitiva e de fácil uso, com apenas alguns cliques já é possível criar um relatório simples.

The screenshot shows the Pentaho CDE (Community Data Explorer) web interface. The browser address bar shows 'localhost:8080/pentaho/Home'. The page has a navigation bar with 'Saiku Analytics' and 'Relatório Vendas'. Below this, there's a section for 'Output Type' set to 'PDF', with 'View Report' and 'Auto-Submit' buttons. The main content area displays a report titled 'Venda Anual Ano 2012' with a table of sales data for 2012. The table has columns for month, seller, department, and sales value. The data is grouped by month, with each month having two rows for different departments (Informática and Ótica). The total sales for each month are listed at the bottom of each group.

nom_mes	nom_vendedor	nom_departamento	val_venda
Abril	Ana Maria	Informática	5.472,00
Abril	Ana Maria	Ótica	3.240,00
Agosto	Ana Maria	Informática	32.230,00
Agosto	Ana Maria	Ótica	47.910,00
Fevereiro	Ana Maria	Informática	19.880,00
Fevereiro	Ana Maria	Ótica	10.910,00
Julho	Ana Maria	Informática	48.190,00
Julho	Ana Maria	Ótica	45.664,00
Junho	Ana Maria	Informática	30.268,00
Junho	Ana Maria	Ótica	37.515,00
Maio	Ana Maria	Informática	8.190,00
Maio	Ana Maria	Ótica	84.294,00

O CDE permite o desenvolvimento e implantação de *dashboards* de forma rápida e eficaz. Abaixo imagem da ferramenta:



Considerações Finais

A intenção deste artigo é proporcionar o seu primeiro contato com a área de Business Intelligence utilizando Pentaho. Procure agora novas dicas e material sobre o assunto, busque treinamento específico para o tema e se aperfeiçoe ainda mais com Pentaho BI.

Acompanhe o BI NA PRÁTICA clicando **AQUI**. Conheça os outros artigos na seção **ARTIGOS**.