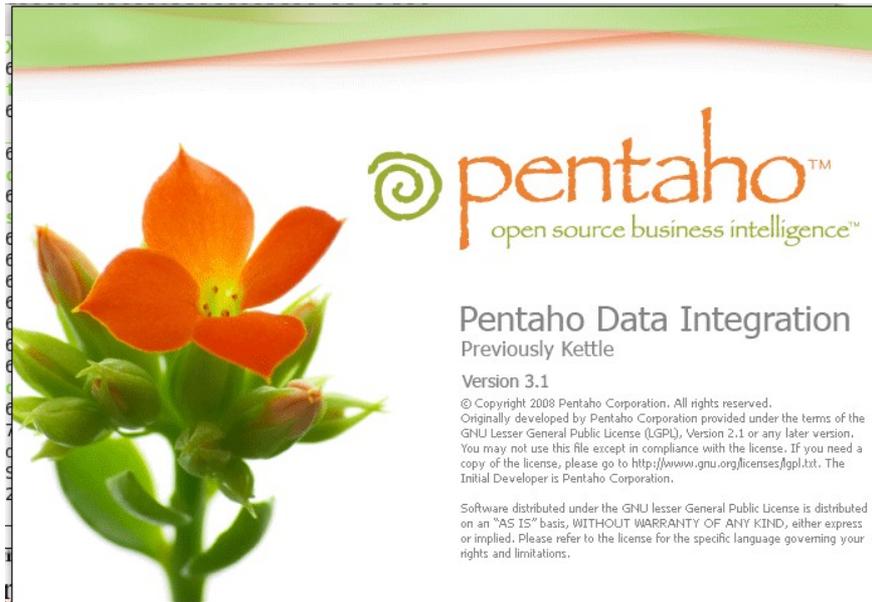


## Pentaho Data Integration (PDI)



O PDI é a ferramenta da plataforma responsável pelo processo de Extração, Transformação e Carga(Load) – ETL, com ele é possível criar processos que automatizam essa tarefa. Nesse momento os dados do modelo transacional serão carregados no modelo dimensional, os dados serão consolidados.

Nas tabelas dimensão são armazenados as chaves naturais, as Pk's do sistema OLTP, bem como as chaves delegada, artificiais SK (Surrogate Key). Podemos concluir que as tabelas de dimensão fazem a ligação do modelo dimensional com o transacional, já na tabela fato as Pk's do OLTP não serão populadas.

Então temos na tabela de dimensão, as PK ligando ao modelo transacional e as SK ligando a tabela Fato.

Pelo PDI você pode desenvolver *Transformação e Jobs*.

### ***Iniciando o PDI***

- 1) acesse o diretório do PDI ( pentaho/pdi-open-3.1.0-826 )
- 2) execute o arquivo de inicialização: spoon.sh  
`sudo ./spoon.sh`



*Ilustração 1: Tela inicial do PDI*

Para entender como o PDI executa uma transformação ou um job, é necessário compreender a lógica utilizada:

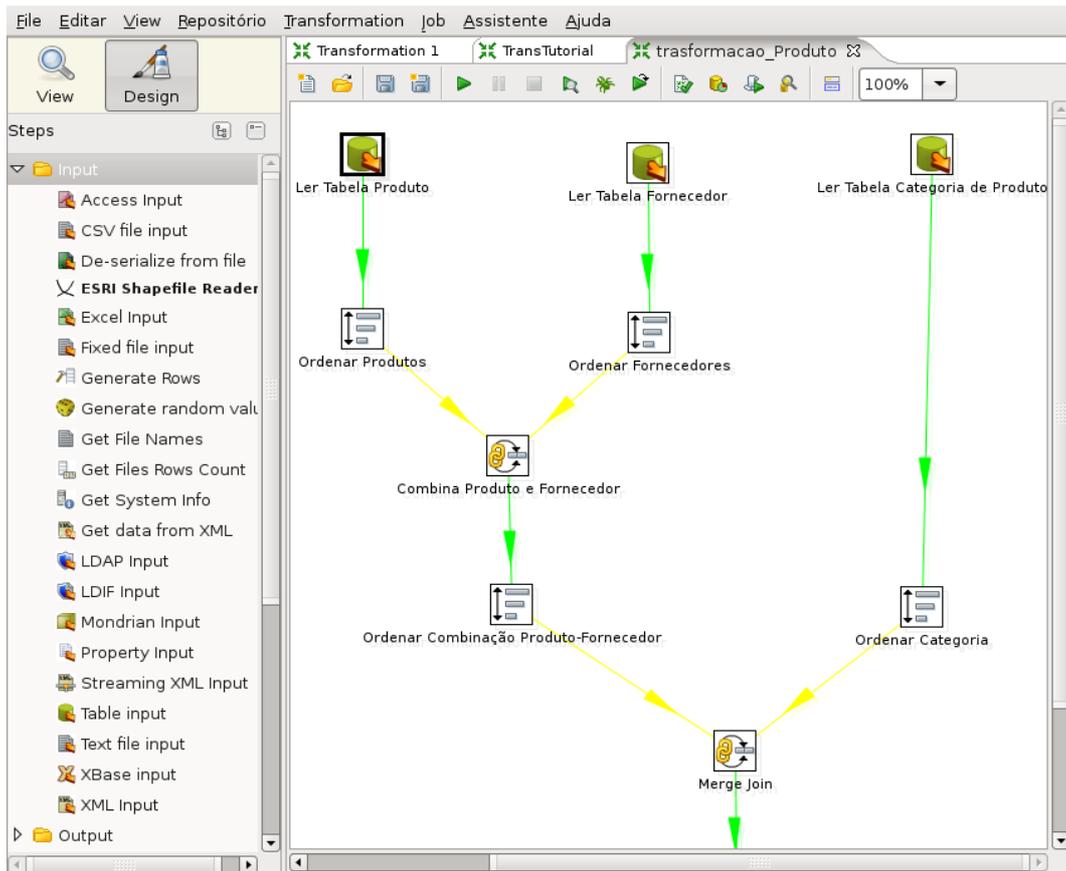
- Os dados fluem entre os passos (steps);
- Os passos são interligados através de ligações (hops);
- Apesar de não ser visível, o fluxo de dados passa pelas ligações, indo de um passo para outro.

Em alguns casos o PDI pode dividir a saída de dados de um passo em duas. Quando isso ocorre, há duas alternativas: os dados são copiados (isto é, fluem em duplicidade pelas duas ligações) ou são distribuídos (uma metade para cada ligação)

## **Transformações**

O objetivo dessa e de qualquer transformação é pegar um dado, sobre esse dado realizar algum ajuste e passar esse dados para outro lugar. A origem e o destino podem ser uma tabela de um banco de dados, uma planilha do excel, um arquivo texto.

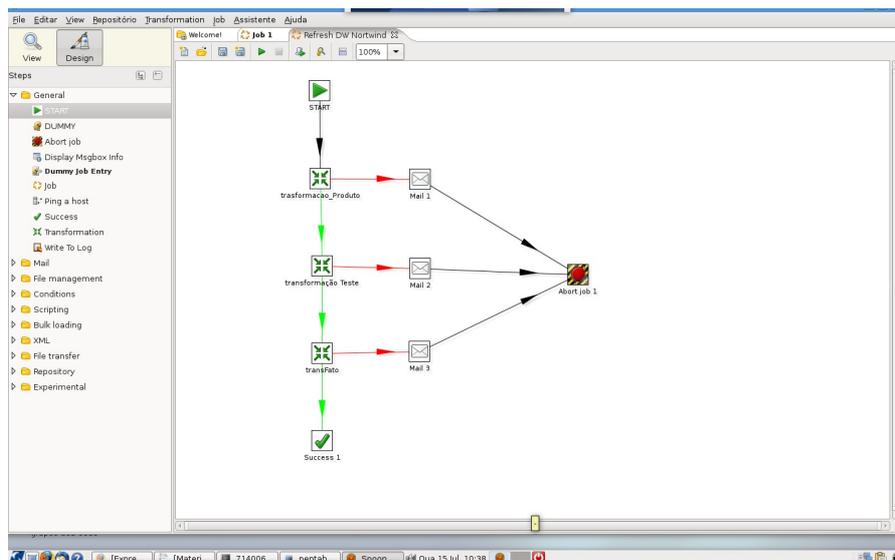
Para iniciar os trabalhos no PDI é necessário a criação de uma conexão como banco de dados(ou qualquer outra origem), essa conexão poderá ser compartilhada entre as transformações.



*Ilustração 2: Exemplo de Transformações*

## Jobs

O job é um recurso do PDI bem útil para automação de rotinas de carga, com ele é possível criar uma sequencia de passos, que podem ser chamadas para transformações existentes, e o mais interessante, possibilita o controle desses passos.



*Ilustração 3: Job do PDI*

As setas que definem o fluxo de execução recebem cores com a seguinte significação:

**Verde** = Siga caso seja Verdadeiro;

**Vermelha** = Siga caso seja Falso;

**Preta** = Siga incondicionalmente.

O PDI oferece uma interface amigável para criação e execução de transformações e jobs, mas é possível via linha de comando executar-los. Dentro do diretório do PDI existem dois aplicativos pan.sh e kitch.sh

pan.sh – utilizado para executar transformações;

kitch.sh – utilizado para executar jobs.

A sintaxe para utilizar esse aplicativos é:

`./pan.sh -file=nomedatransformacao.ktr`

`./kitch.sh -file=nomedojob.kjb`

Não seria muito interessante abrir o shell para executar essas chamadas, uma vez que existe uma interface para isso, mas esse recurso permite criar agendamentos via crontab. Para editar o crontab utilize a sintaxe: *crontab -e*

*Espero te ajudado!*

**Bruno Jackson Iaccino Coelho**

E-mail: [bruno.iaccino@gmail.com](mailto:bruno.iaccino@gmail.com)