

01 { [Criando um Data Warehouse com Pentaho]



1
2
3
4
5
6
7
8
9
10
11
12
13
14

}

Por que usar o **Pentaho**?



1
2
3
4
5
6
7
8
9
10
11
12
13
14

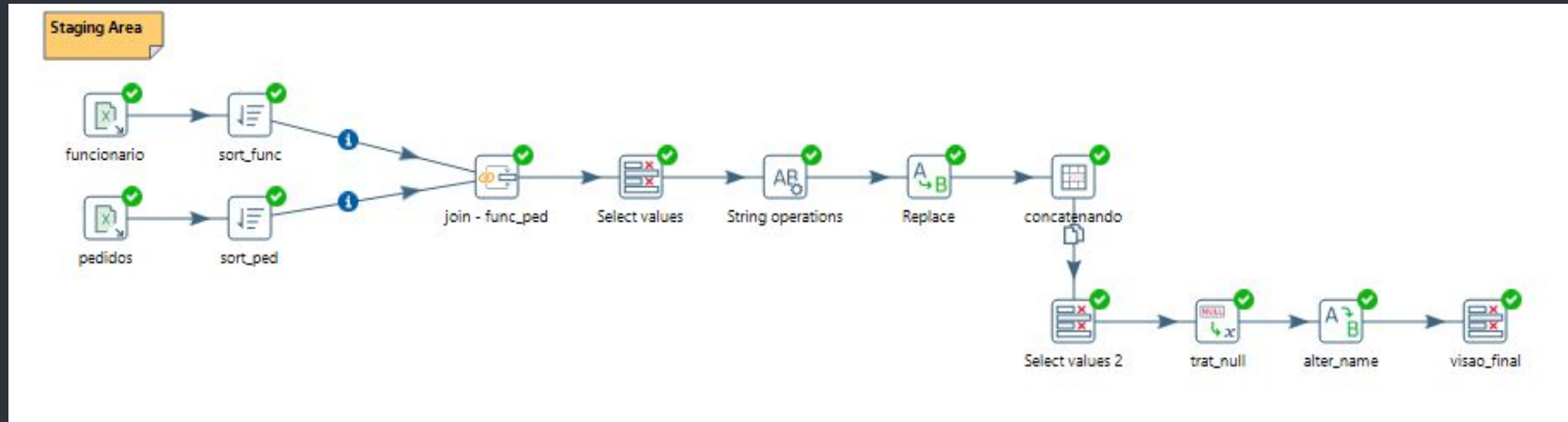
O Pentaho fornece dados com muita facilidade e faz com que seus relatórios sejam mais minuciosos e detalhistas.

Utilizando-os de maneira correta, você compreende melhor como sua gestão funciona e pode alocar recursos com mais eficiência.

}



Realizando ETL/ELT com planilhas



[Preparando o Ambiente]

Windows

- Instalar docker - utilizei a versão 4.3

No terminal:

Instalar no docker o Postgres

```
docker pull postgres
```

```
docker run --name nome-do-docker -e POSTGRES_PASSWORD=suasenha -d  
postgres
```

Exibir o postgres para acesso e criação de local para armazenar:

```
docker run -p 5432:5432 -v /tmp/database:/var/lib/postgresql/data -e  
POSTGRES_PASSWORD=suasenha postgres
```

}



[Preparando o Ambiente]

Windows

- Instalar no docker o Mysql

```
docker run --name nome-do-docker -e MYSQL_ROOT_PASSWORD=suasenha -d mysql:8.0.33
```

- Exibir o postegres para acesso e criação de local para armazenar:

```
docker run --name nome-do-docker -v /tmp/database:/var/lib/mysql/data -e  
MYSQL_ROOT_PASSWORD=suasenha -d mysql:8.0.33
```

Executar Script DDL para alimentar o BD

- Instalar o Java:

```
java jdk e jre - vrs: 8u311
```

- Criar Variavel de ambiente:

pesquisar no win → var → nova variavel → selecionar a pasta do jdk

- Instalar o Pentaho - PDI:

```
https://sourceforge.net/projects/pentaho/files/latest/download - vrs:  
pdi-ce-9.3.0.0-428
```

basta "deszipar" o arquivo no repositorio central do windows c:



[Preparando o Ambiente]

Windows

- Configs:

Colar o drive do banco a ser usado na pasta
C:\pdi-ce-9.3.0.0-428\data-integration\lib

No meu caso utilizei o Drive mssql-jdbc-9.2.0.jre8 e o autenticador
mssql-jdbc_auth-9.2.0.x64.dll(colocar ele na pasta:
C:\pdi-ce-9.3.0.0-428\data-integration\libswt\win64)

Enviar como atalho para area de trabalho o arquivo .bat: "Spoon",
renomea-lo para "Pentaho".



[Hands ON]

Pentaho

*CRIAR AS PASTAS POR ÁREAS:

1º - Apresentar a instalação e disponibilizar o passo a passo;

2º - Mostrar como Conectar os bancos de dados e Planilhas;

Extract

2.1 - table_input(BD) → excel(salvar pelo nome e abrir content) - write

2.2 - excel - imputar o excel - UTILIZAR PEDIDO E FUNCIONARIO

2.4 - documentar(botão direito)

2.3 - Bancos de dados(apresentar tb a questão dos drives)



[Hands ON]

Pentaho

3º - Joins

3.1 - Sempre que for executar o join fazer o sort(verificar a ordenação)

3.2 - Exemplo



[Hands ON]

Pentaho

4º - Tratamento dos dados

4.1 - tratar as datas - se utiliza o select values e não fazer direto.

4.2 - tratar os tipos de dados

4.3 - tratar strings (String operation), endereço;

4.4 - Mais tratamentos de dados (Replace in string);

4.5 - Concatenar(concat fields)

4.6 - Tratando nulos (null)

4.7 - Alterando nomes (value mapper)



[Hands ON]

Pentaho

Parte 3

5º - Logs*

5.1 - clicar na parte branca e ir na aba log(SALVAR INFORMAÇÕES ENTRE BANCOS)

6º - Dimensões e Fatos

6.1 - Criar Dim Full - usar o truncate para apagar tudo e criar novamente

6.2 - Criar o arquivo de erro(text file output)

6.3 - simular um erro e alterar o total de caracter em uma coluna para criar.



[Hands ON]

Pentaho

Parte 4

6º - Carga Incremental - armazenando o historico de atualização

7.1 - dimension lookup/update → startdate data da transfor

7.2 - incrementar:

```
update north.customers
set city = 'Joao Pessoa'
WHERE customer_id = 'WOLZA'
```

```
INSERT INTO cliente (nome, email, telefone, endereco) VALUES
('João da Silva', 'joao.silva@gmail.com', '(11) 9999-9999', 'Rua A, 123')
```

}



[Hands ON]

Pentaho

Parte 5

7.3 – usar a mesma base(puxei do sqlserver – produto)

7.4 – usar o short para ordenar

7.5 – usar o merge rows (diff)

De um lado eu vou ter as chaves e do outro os valores, primeira vez que rodar vai ficar igual;

criar o merge com o Dummy e depois inserir um novo item na tabela;

}



[Hands ON]

Pentaho

```
USE [Traders]
```

```
GO
```

```
INSERT INTO [dbo].[Produtos]
```

```
    ([NomeDoProduto]
```

```
    ,[CodigoDoFornecedor]
```

```
    ,[CodigoDaCategoria]
```

```
    ,[QuantidadePorUnidade]
```

```
    ,[PrecoUnitario]
```

```
    ,[UnidadesEmEstoque]
```

```
    ,[UnidadesPedidas]
```

```
    ,[NivelDeReposicao]
```

```
    ,[Descontinuado])
```

```
VALUES
```

```
    ('JUMANJY', 1, 1, '9 caixas x 20 peças', '250,00', 40, 10, 5, '0')
```

```
GO
```



uniesp

Centro Universitário