

Turma DA-18

# APRESENTAÇÃO DE BRIEFING DO PROJETO

Encerramento da Unidade 02 - ETL

**Facilitadora:**

Nayara Valevskii

**Data Apresentação:**

15/03/2025 - para participar do sorteio, até dia 22/03 prazo final  
(prazo extra de entrega).

Feedback envio até fim do mês individual.



# TÓPICOS

- O2** Vai ser entregue em equipe?
- O3** Visão Geral
- O4** Sobre a Base de Dados
- O5** Objetivos e Metas
- O6** Atividades
- O7** Dicas e Motivação

# INDIVIDUAL

- O Trabalho vai ser um exercício individual, nada complexo.
- O Objetivo é avaliar a capacidade de construção dos pipelines de stage e datawarehouse, proporcionando um feedback de forma individualizada.



# VISÃO GERAL

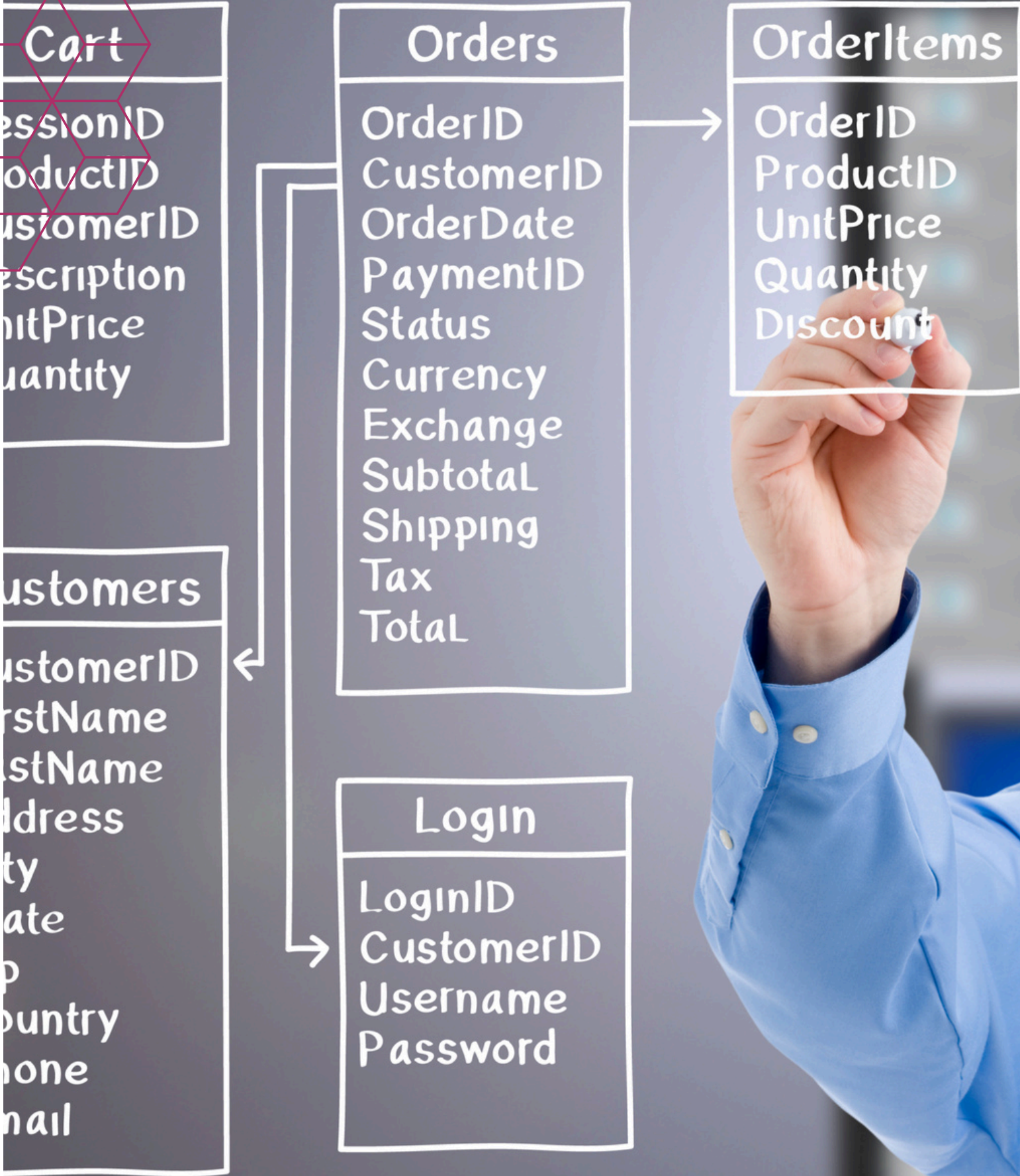
O objetivo deste projeto é construir um ambiente de análise de dados a partir de arquivos CSVs, utilizando as técnicas de ETL aprendidas em sala de aula para criar um Data Warehouse eficaz. Além disso, o projeto visa desenvolver análises que servirão como base para tomadas de decisões estratégicas para o negócio.





# BASE DE DADOS

A área de negócio escolhida para o projeto foi uma base de RH.





# ESTRUTURA DOS DADOS

Para facilitar a construção do projeto, dentro do drive vai conter:

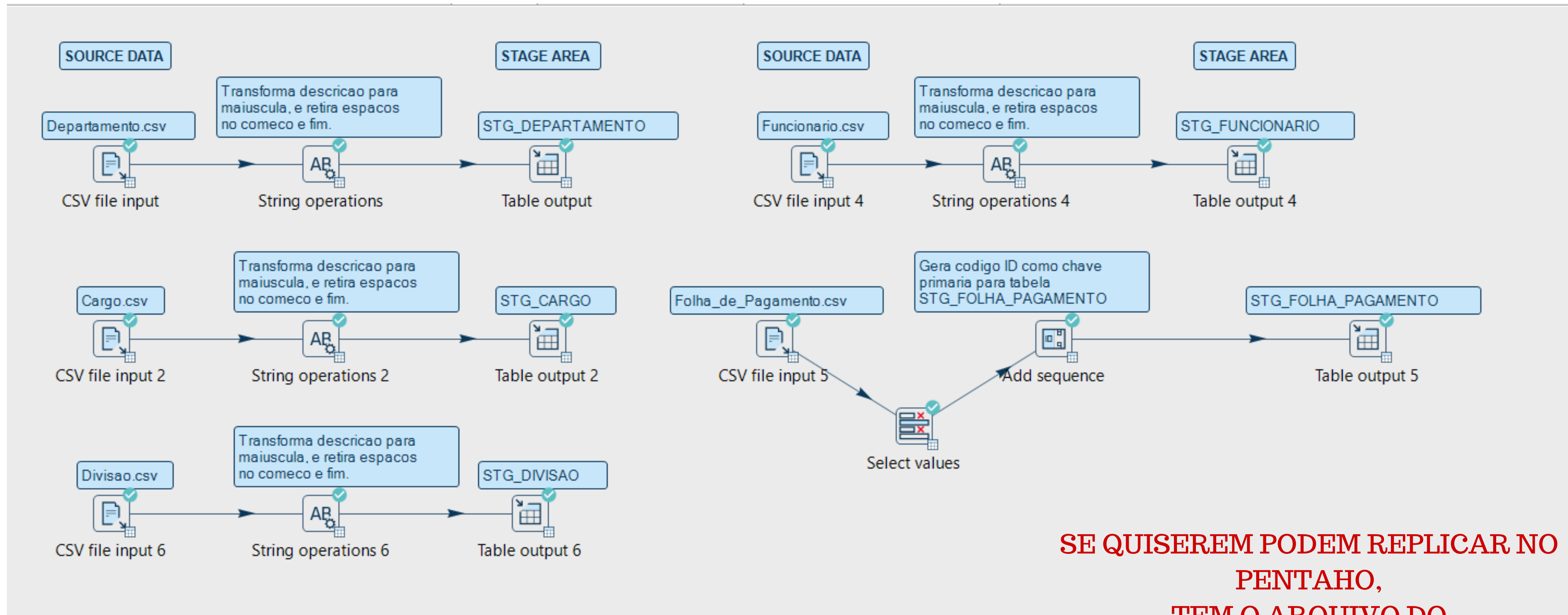
- Datasets do projeto (arquivos csv's).
- Scripts SQL com a criação das tabelas do Stage e Datawarehouse.
- Backup do banco de dados Stage (só com as tabelas vazias e com as tabelas preenchidas) e banco de dados Datawarehouse com as tabelas (sem os dados, precisa processar os dados no pipeline).
- Imagem dos fluxos de pipelines do Stage e 1 exemplo de fluxo de pipeline de uma dimensão.
- Arquivos dos fluxos no Apache Hop (Stage apenas como exemplo).
- Arquivos dos fluxos no Pentaho (2 dimensões).
- LINK: <https://drive.google.com/drive/folders/1oIFgkBkgVmVyIr18Q54Q0Y8KkNkpSyQ?usp=sharing>



**Acesse a base  
de dados AQUI**

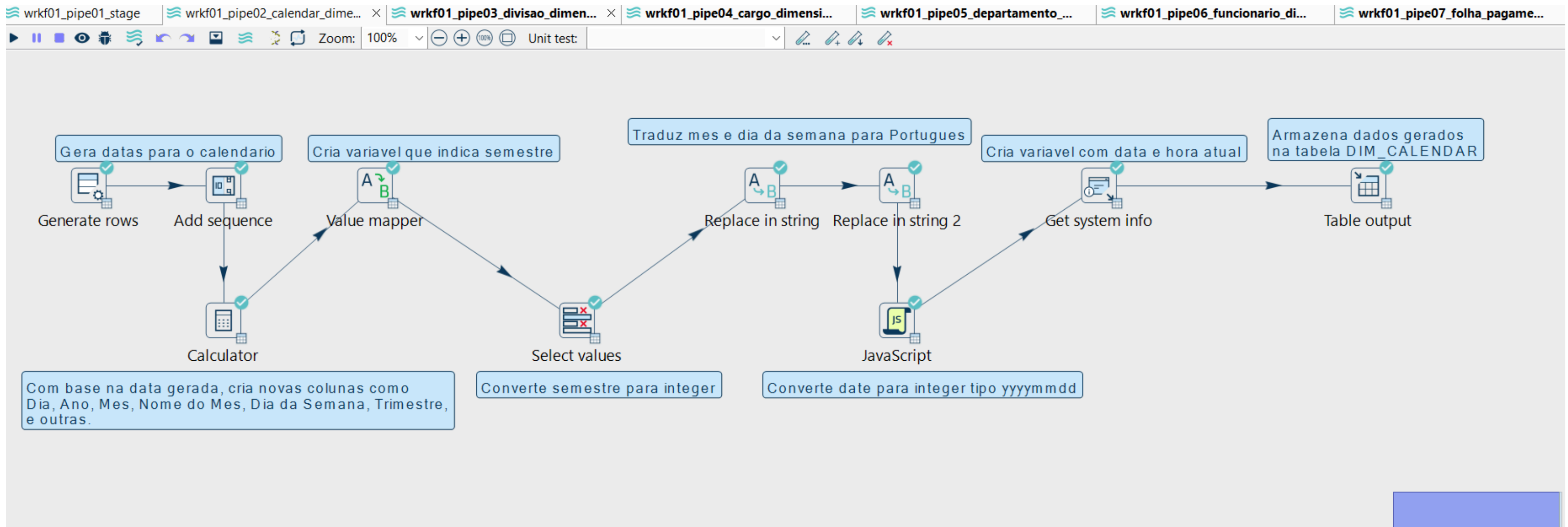


# PIPELINE STAGE



SE QUISEREM PODEM REPLICAR NO  
PENTAHO,  
TEM O ARQUIVO DO  
HOPE SALVO

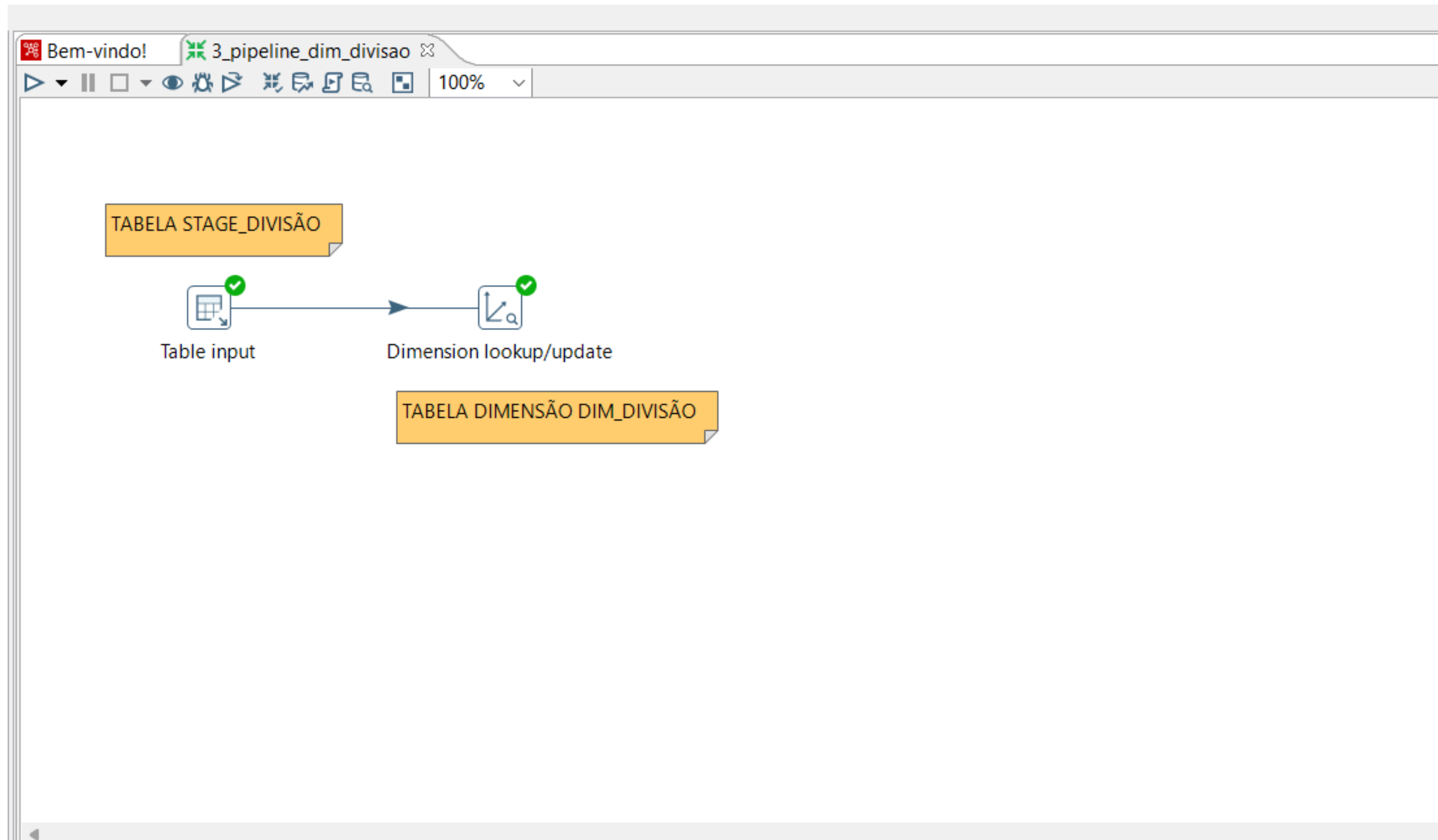
# PIPELINE DIMENSÃO CALENDÁRIO



NO BACKUP DO BANCO JÁ ESTÁ COM A DIMENSÃO CALENDÁRIO PREENCHIDA, MAS DEIXEI O EXEMPLO SALVO DO PIPELINE DO HOPE CASO ALGUÉM QUEIRA UTILIZAR.

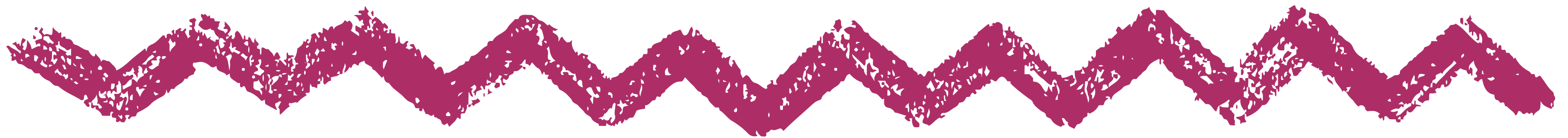


# PIPELINE DIMENSÃO DIVISÃO

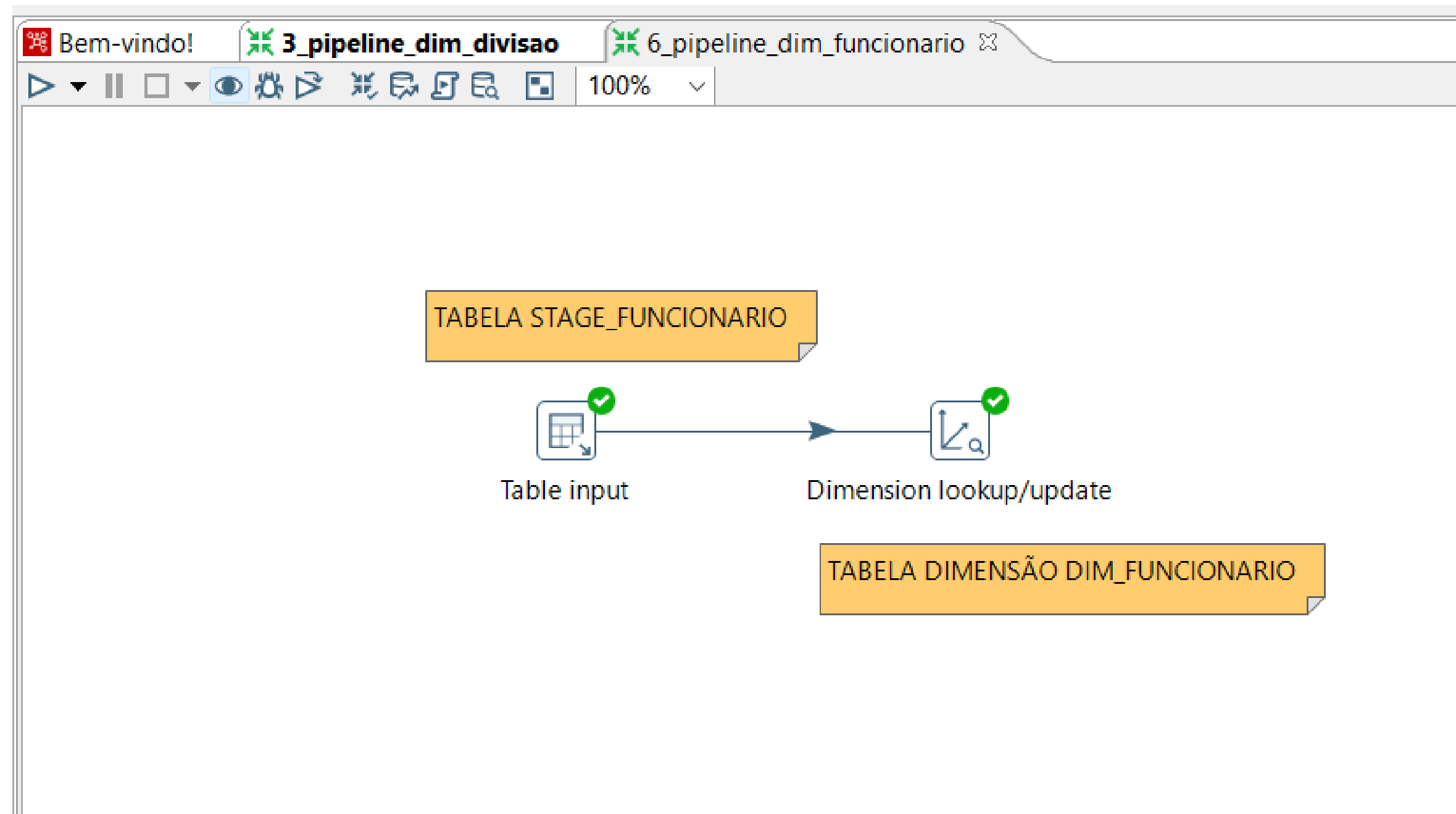




**ANTES DE PROCESSAR A  
DIMENSÃO FUNCIONÁRIO,  
PRECISAM CONSTRUIR OS  
PIPELINES E PROCESSAR A  
DIMENSÃO CARGO E DEPOIS  
DIMENSÃO DEPARTAMENTO.**



# PIPELINE DIMENSÃO FUNCIONÁRIO







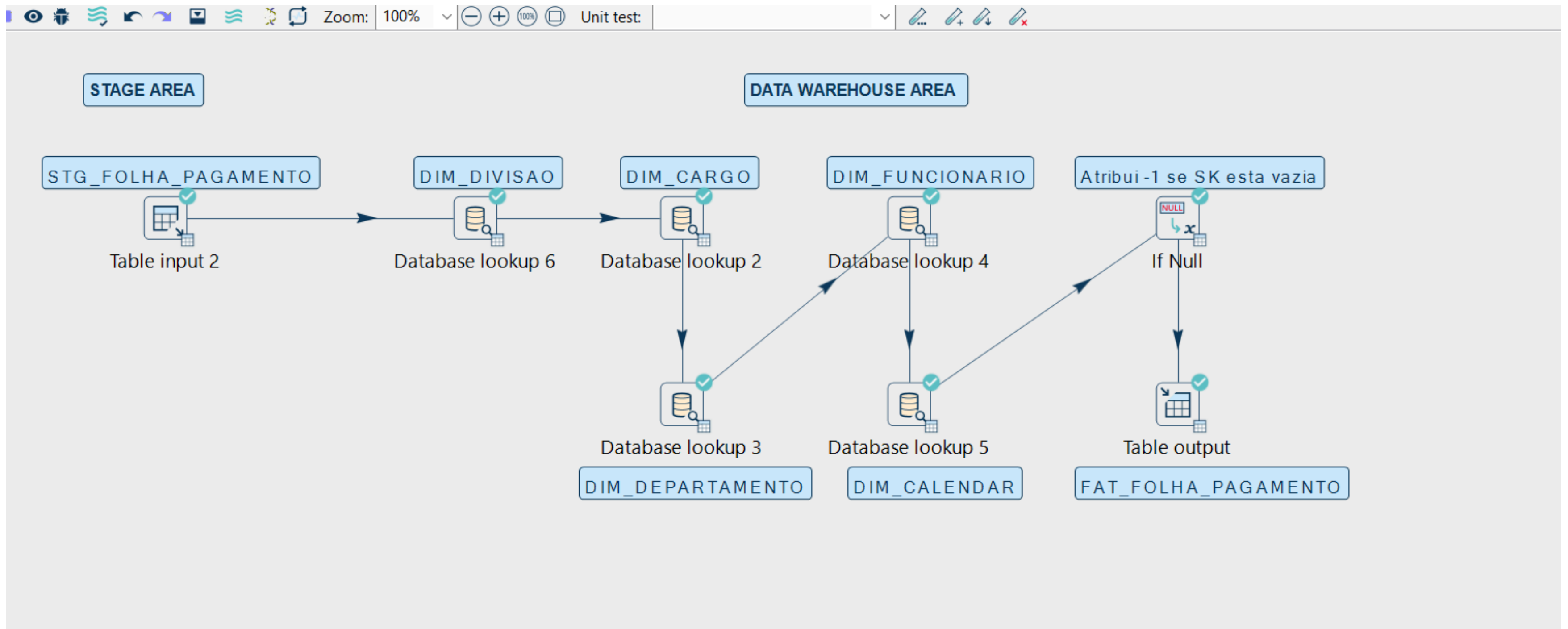
# **SOBRE OS PIPELINES PENDENTES, QUE PRECISAM PARA O PROJETO!**

## **DICAS!!!!**

- Falta fazer o Pipeline para a dimensão\_cargo e dimensão\_departamento, a lógica é a mesma da dimensão divisão.
- Para o pipeline da folha de pagamento, vou deixar o print do pipeline para seguirem. Lembre-se que no lookup precisa ir as datas de inicio e fim? Recordam? E o lookup entra para buscar os ids das dimensões. Vou deixar o print do scrip do if null apenas... os demais vocês conseguem desenrolar, certo?

# PIPELINE FATO FOLHA

## EXEMPLO



[illegible]



# OBJETIVOS E METAS

Para o projeto cumprir os objetivos e metas principais de entrega.

**#1**

Criar os pipelines das dimensões pendentes e a fato folha.

**#2**

Criar um Job para processar o Stage e o DW.

**#3**

Criar análises detalhadas que respondam aos requisitos do tópico "Atividades", oferecendo insights relevantes em SQL.

# ANÁLISES OBRIGATÓRIAS

Apresentar algumas análises obrigatórias para o projeto, em SQL.. ou seja as queries que respondam a pergunta.

```

21 SELECT DISTINCT pp.LastName, pp.
22 FROM Person.Person pp JOIN Human
23 ON e.BusinessEntityID = pp.Busi
24 (SELECT SalesPersonID
25 FROM Sales.SalesOrderHeader
26 WHERE SalesOrderID IN
27 (SELECT SalesOrderID
28 FROM Sales.SalesOrderDetail
29 WHERE ProductID IN

```

# ATIVIDADES

Restaurar o banco de dados **Stage e DataWarehouse**.

Processar os pipelines já entregues e construir os pendentes.

Criar análises SQL, para as perguntas abaixo:

## KPIS

- TOTAL DE FUNCIONÁRIOS POR DEPARTAMENTO;
- MÉDIA SALARIAL POR CARGO;
- MÉDIA TEMPO DE EMPRESA;
- CUSTO TOTAL FOLHA DE PAGAMENTO;

## ANÁLISES

- FUNCIONÁRIOS COM MAIOR TEMPO DE EMPRESA;
- DISTRIBUIÇÃO SALARIAL POR CARGO;
- EVOLUÇÃO DO CUSTO DE FOLHA DE PAGAMENTO;

**ENTREGA EXTRA  
FAZER AS ANÁLISES NO POWERBI**





# DICAS E MOTIVAÇÃO



Lembre-se que estamos  
entre amigos, não precisa  
ficar tímido(a).

Aproveitem a atividade  
individual para identificar os  
pontos de dificuldades.

Use sua criatividade!

Não esqueçam de aplicar as  
etapas que aplicamos em  
um projeto de Dados!

Se é motivação que você  
precisa, aqui vai uma: VAI  
TER SORTEIO!

Para participar do sorteio,  
entregar o projeto final até  
dia 14/03.

E SIMBORA PARA UNIDADE  
3!

# OBRIGADA!

Enviar a atividade até o dia 14/03, para participar do sorteio (olha a motivação), mas estou postergando o prazo de entrega até dia 22/03, para que recebam o feedback personalizado e dar tempo fazer o BI.

e-mail: [nayara.valevskii@gmail.com](mailto:nayara.valevskii@gmail.com)

Boa sorte!

