Documentação Completa - refined.d\_embarcacao\_farol

## Tabela: refined.d\_embarcacao\_farol

### Objetivo:  
Tabela da dimensão de pessoas do indicador Farol.

### Fontes de Dados  
  
| Origem | Descrição |  
|---------------------------------------|----------------------------------------------------------------|  
| trusted.tb\_tab\_pessoa\_farol | Tabela com dados cadastrais de pessoas físicas e jurídicas. Funcionários. |

### Histórico de alterações  
  
| Data | Desenvolvido por | Modificações |  
|------------|------------------|-----------------------|  
| 21/05/2025 | Michel Santana | Criação do notebook |

# Código do Notebook

## Bloco de Código 1

# Importa e executa o notebook `ingestion\_function`, localizado em `../00\_config/`.  
# O comando `%run` carrega todas as funções, variáveis e configurações definidas no notebook referenciado  
# para o ambiente atual. Isso permite reutilizar lógica comum, como funções de ingestão de dados, sem duplicação de código.  
# Útil para centralizar rotinas reutilizáveis e manter notebooks modulares e organizados.

## Bloco de Código 2

%run ../00\_config/ingestion\_function

## Bloco de Código 3

container\_target = 'refined'  
directory = 'farol'  
table\_name = 'd\_embarcacao\_farol'  
delta\_table\_name = f'{environment}.{container\_target}.{table\_name}'  
delta\_file = f"abfss://{container\_target}@{data\_lake\_name}.dfs.core.windows.net/{directory}/{table\_name}/"  
comment\_delta\_table = 'Tabela de dimensão de Pessoas...'  
print(f'delta\_table\_name = {table\_name}')  
print(f'delta\_file = {delta\_file}')

## Bloco de Código 4

"""  
Cria um widget interativo chamado `reprocessar` para controle da carga da dimensão.  
- Exibe um dropdown com as opções "True" e "False".  
- A variável `reprocessar` será `True` apenas se o usuário selecionar essa opção no notebook.  
Objetivo: permitir que o usuário escolha, de forma interativa, se a carga será completa (`overwrite`) ou incremental (`merge`).  
"""

## Bloco de Código 5

dbutils.widgets.dropdown("reprocessar", "False", ["True", "False"], "Reprocessar dimensão?")  
reprocessar = dbutils.widgets.get("reprocessar") == "True"

## Bloco de Código 6

spark.sql(f"""  
CREATE TABLE IF NOT EXISTS {delta\_table\_name} (  
 sk\_embarcacao BIGINT GENERATED ALWAYS AS IDENTITY (START WITH 1 INCREMENT BY 1),  
 sequencia\_embarcacao INT,  
 descricao\_embracacao STRING,  
 numero\_registro\_capitania STRING,  
 numero\_registro\_mpa STRING,  
 informacao\_embarcacao\_1 STRING,  
 informacao\_embarcacao\_2 STRING,  
 insert\_date TIMESTAMP,  
 update\_date TIMESTAMP  
)  
USING DELTA  
LOCATION '{delta\_file}'  
COMMENT '{comment\_delta\_table}';  
""")

## Bloco de Código 7

spark.sql(f"""  
select  
 seq\_embarcacao as sequencia\_embarcacao,  
 des\_embarcacao as descricao\_embracacao,  
 num\_registro\_capitania as numero\_registro\_capitania,  
 num\_registro\_mpa as numero\_registro\_mpa,  
 inf\_embarcacao\_1 as informacao\_embarcacao\_1,  
 inf\_embarcacao\_2 as informacao\_embarcacao\_2  
from {environment}.trusted.tb\_tab\_pessoa\_embarcacao\_farol  
where 1=1  
""").createOrReplaceTempView("source\_df")

## Bloco de Código 8

if reprocessar:  
 source\_df = source\_df.withColumn("insert\_date", lit(current\_timestamp() ) )  
 source\_df.write.format("delta")\  
 .mode("overwrite")\  
 .saveAsTable(f"{delta\_table\_name}", path=f"{delta\_file}")  
 print(f"Carga overwrite realizada com sucesso! {delta\_table\_name}")  
else:  
 print("Realizando carga em Merge..")  
 spark.sql(f"""  
 MERGE INTO {delta\_table\_name} AS target  
 USING source\_df AS source  
 ON target.sequencia\_embarcacao = source.sequencia\_embarcacao  
 WHEN MATCHED THEN  
 UPDATE SET  
 target.descricao\_embracacao = source.descricao\_embracacao,  
 target.numero\_registro\_capitania = source.numero\_registro\_capitania,  
 target.numero\_registro\_mpa = source.numero\_registro\_mpa,  
 target.informacao\_embarcacao\_1 = source.informacao\_embarcacao\_1,  
 target.informacao\_embarcacao\_2 = source.informacao\_embarcacao\_2,  
 target.update\_date = current\_timestamp()  
 WHEN NOT MATCHED THEN  
 INSERT (  
 sequencia\_embarcacao,  
 descricao\_embracacao,  
 numero\_registro\_capitania,  
 numero\_registro\_mpa,  
 informacao\_embarcacao\_1,  
 informacao\_embarcacao\_2,  
 insert\_date  
 )  
 VALUES (  
 source.sequencia\_embarcacao,  
 source.descricao\_embracacao,  
 source.numero\_registro\_capitania,  
 source.numero\_registro\_mpa,  
 source.informacao\_embarcacao\_1,  
 source.informacao\_embarcacao\_2,  
 current\_timestamp()  
 );  
 """)