

amcom

inspired by tech, powered by people

Sumário

1.	Visão Geral.....	3
2.	Origem.....	3
2.1	Tabela.....	3
2.2	Qualidade origem.....	4
3.	Raw.....	4
3.1	Tabela.....	5
4.	Trusted	5
4.1	Tabela.....	5
5.	Refined	6
5.1	Tabela.....	6
6.	QUALIDADE.....	6
7.	DETALHAMENTO	7

1. Visão Geral

Este documento apresenta a arquitetura de dados da tabela **Facas**, especificando as fontes de origem e as estruturas existentes em cada camada do pipeline de tratamento e processamento de dados.

Facas:

“Cadastro de Facas: Facas são ferramentais de corte necessários para conformação total das caixas tipo 'C/V' (corte & vinco). Nas caixas do tipo 'Flexo' não são necessárias, porém são utilizadas para efetuar pequenos detalhes quando solicitados.”.

2. Origem

Banco: SQL Server

Base de dados: DBO

Tabela: Facas

Select simples: `select * from dbo.Facas`

2.1 Tabela

A tabela abaixo contempla a estrutura atual da fonte Facas com a definição de tipo de dado correspondente, onde:

- **Coluna:**
Representa o nome de cada coluna disponível na tabela original.
- **Tipo:**
Apresenta qual o tipo de dado da coluna relacionada, exemplo: int, varchar, etc.
- **Necessária:**
Campo preenchido pela ADAMI, onde informa se a coluna relacionada será necessária para execução do projeto.
- **Ação Sugerida:**
Sugestão preenchida pela AmCom, onde informa se a coluna relacionada deveria ser mantida ou excluída. Para isto foram considerados os seguintes critérios de exclusão:

Muitos Nulos: > 90% de valores nulos
Valor Único: Coluna possui apenas 1 valor único
Muitos zeros: > 80% de valores zero (para colunas numéricas)
Strings vazias: > 80% de strings vazias (para colunas de texto)

Colunas que não atendem a esses critérios são mantidas.
- **Observação:**
Observações relacionadas aos dados e incluídas pela AmCom e/ou ADAMI.

Coluna	Tipo	Necessária?	Ação Sugerida (AmCom)	Observação
CodFaca	varchar	Sim	Manter	
Tipo	int	Não	Manter	
DescrTipoFaca	varchar	Não	Manter	
Fornecedor	varchar	Não	Manter	
Modulo	varchar	Não	Manter	
Arquivo	varchar	Não	Manter	
Largura	int	Não	Excluir	
Comprimento	int	Não	Excluir	
PecasPorFaca	int	Não	Manter	
Localizacao	varchar	Não	Excluir	
Figura	varchar	Não	Manter	
Status	int	Sim	Manter	1-Ativ ; 2-Susp ; 5-Desm
Preco	float	Não	Manter	
Complamina	int	Sim	Manter	Quantidade linear de Lamina contida na faca (mm)
IDCliente	int	Não	Manter	
Obs	varchar	Não	Manter	
CodigoERP	varchar	Não	Excluir	
ChaveEstadoFerramenta	int	Não	Excluir	
DesativadoSN	bit	Sim	Manter	"=0 p/ Ativas"
RefugoCliente	float	Não	Excluir	
DataCriacaoRegistro	datetime	Não	Manter	

2.2 Qualidade origem

Com o intuito de compreender e gerenciar a qualidade dos dados disponíveis na origem, foi conduzido um trabalho de validação detalhado. Cada coluna foi avaliada individualmente quanto à completude, consistência, padronização e presença de valores inválidos ou ausentes. Esse processo permitiu identificar inconsistências e definir as regras de limpeza e transformação necessárias para garantir a confiabilidade dos dados nas etapas subsequentes do pipeline.

O artefato correspondente a esta validação está disponível em:

https://github.com/Adami-S-A/Projeto_IA_AMCOM/tree/main/project_data_science/docs/data_quality

3. Raw

A camada *Raw* é a primeira etapa do pipeline de dados, onde as informações são armazenadas exatamente como são recebidas da fonte, sem qualquer tipo de transformação ou limpeza. Essa camada preserva o dado em seu formato original, incluindo possíveis inconsistências, duplicidades ou valores faltantes. Seu principal objetivo é servir como backup confiável do dado bruto, garantindo rastreabilidade e permitindo auditoria e reconstrução de dados caso seja necessário.

3.1 Tabela

A tabela abaixo contempla a estrutura disponibilizada na camada *raw* para a tabela Facas.

Nome	Tipo
CODFACA	VARCHAR2(30)
TIPO	NUMBER
DESCRTIPOFACA	VARCHAR2(30)
FORNECEDOR	VARCHAR2(25)
MODULO	VARCHAR2(50)
ARQUIVO	VARCHAR2(50)
LARGURA	NUMBER
COMPRIMENTO	NUMBER
PECASPORFACA	NUMBER
LOCALIZACAO	VARCHAR2(50)
FIGURA	VARCHAR2(99)
STATUS	NUMBER
PRECO	BINARY_DOUBLE
COMPLAMINA	NUMBER
IDCLIENTE	NUMBER
OBS	VARCHAR2(1000)
CODIGOERP	VARCHAR2(30)
CHAVEESTADOFERRAMENTA	NUMBER
DESATIVADOSN	CHAR(1)
REFUGOCLIENTE	BINARY_DOUBLE
DATACRIACAOREGISTRO	DATE

4. Trusted

A camada Trusted é composta por dados que passaram por processos de validação e limpeza, tornando-os consistentes e confiáveis para análises. Nessa camada, são aplicadas regras de qualidade, padronização de formatos, remoção de duplicidades e tratamento de valores nulos críticos. Todas as colunas sinalizadas com a opção de exclusão durante a análise de entendimento foram desconsideradas.

4.1 Tabela

A tabela abaixo contempla a estrutura disponibilizada na camada *trusted* para a tabela Facas.

Name	Type
CODFACA	VARCHAR2(30)
TIPO	NUMBER
DESCRTIPOFACA	VARCHAR2(30)
FORNECEDOR	VARCHAR2(25)
MODULO	VARCHAR2(50)
ARQUIVO	VARCHAR2(50)

PECASPORFACA	NUMBER
FIGURA	VARCHAR2(99)
STATUS	NUMBER
PRECO	BINARY_DOUBLE
COMPLAMINA	NUMBER
IDCLIENTE	NUMBER
OBS	VARCHAR2(1000)
DESATIVADOSN	CHAR(1)
DATACRIACAOREGISTRO	TIMESTAMP(6)

5. Refined

A camada Refined é a etapa final do pipeline, onde os dados são transformados e otimizados para consumo analítico avançado. Inclui cálculos, agregações, junções entre múltiplas fontes e criação de indicadores ou métricas derivadas. Nesta camada, os dados estão modelados para suportar dashboards, relatórios estratégicos, machine learning e tomada de decisão de negócio, proporcionando insights de alto valor a partir de informações confiáveis e enriquecidas.

5.1 Tabela

A tabela abaixo contempla a estrutura disponibilizada na camada *refined* para a tabela Facas.

Name	Type
CD_FACA	VARCHAR2(30)
ST_STATUS	VARCHAR2(5)
VL_COMPLAMINA	NUMBER
FL_DESATIVADOSN	VARCHAR2(1)

6. QUALIDADE

Foi realizada a análise da qualidade dos dados disponíveis na camada *refined*.

Name	Qualidade verificada
CD_FACA	Valores NULL: 0 (0%) Tamanho: 18
ST_STATUS	Valores NULL: 0 (0%) Tamanho: 1
VL_COMPLAMINA	Valores NULL: 0 (0%) Valores com decimais: 0 Valores NaN/Infinito: 0 Mínimo: 0, Máximo: 117050
FL_DESATIVADOSN	Valores NULL: 0 (0%)

7. DETALHAMENTO

Para mais detalhes sobre esta tabela, consulte o documento:
project_data_science/docs/Analise/dicionarioDados.html