# Dicionário de Dados - Camada Silver

Este documento descreve a estrutura e os metadados das tabelas na Camada Silver do Data Lake. A Camada Silver contém dados limpos, com tipos corrigidos, e estruturados, prontos para serem consumidos pelos pipelines de agregação da Camada Gold. Os dados nesta camada são armazenados no formato **Delta Lake**.

## 1. Tabela do Pipeline Histórico (Batch)

### Tabela: posicoes\_onibus

* **Localização:** s3a://silver/posicoes\_onibus
* **Formato:** Delta Lake
* **Particionamento:** Os dados são particionados por ano, mes e dia para otimizar as leituras do pipeline de lote (batch), permitindo que o Spark leia apenas as pastas necessárias para uma determinada hora de processamento.
* **Atualização:** A cada hora, orquestrada pela DAG bronze\_to\_silver do Airflow.
* **Propósito:** Servir como a fonte de dados limpa e histórica para o cálculo da fato\_operacao\_linhas\_hora.

#### Estrutura das Colunas

| **Nome da Coluna** | **Tipo de Dado (Spark)** | **Descrição** | **Origem (Camada Bronze)** |
| --- | --- | --- | --- |
| letreiro\_linha | StringType | Identificador público da linha (ex: "8000-10"). | l.c |
| codigo\_linha | LongType | Código interno/oficial da linha. | l.cl |
| prefixo\_onibus | LongType | Identificador único de cada ônibus. | vs.p |
| acessivel | BooleanType | true se o ônibus é adaptado para acessibilidade. | vs.a |
| timestamp\_captura | TimestampType | Data e hora exata (em UTC) em que a posição foi registrada. | vs.ta |
| data\_processamento | TimestampType | Data e hora em que o registro foi processado pelo Spark. | Gerado pelo Spark |
| ano | IntegerType | **(Coluna de Partição)** Ano extraído do timestamp\_captura. | Gerado pelo Spark |
| mes | IntegerType | **(Coluna de Partição)** Mês extraído do timestamp\_captura. | Gerado pelo Spark |
| dia | IntegerType | **(Coluna de Partição)** Dia extraído do timestamp\_captura. | Gerado pelo Spark |

## 2. Tabela do Pipeline de Tempo Real (Streaming)

### Tabela: posicoes\_onibus\_streaming

* **Localização:** s3a://silver/posicoes\_onibus\_streaming
* **Formato:** Delta Lake
* **Particionamento:** Os dados são particionados por ano, mes, dia e hora para permitir que os processos de streaming subsequentes leiam apenas os dados mais recentes de forma eficiente, evitando a leitura completa da tabela.
* **Atualização:** Em micro-lotes (a cada 2 minutos), pelo processo bronze\_to\_silver\_streaming.
* **Propósito:** Servir como a camada intermediária para os pipelines de streaming da camada Gold. Contém os dados de geolocalização necessários para alimentar os KPIs operacionais em tempo real (mapa, velocidade, etc.).

#### Estrutura das Colunas

| **Nome da Coluna** | **Tipo de Dado (Spark)** | **Descrição** | **Origem (Camada Bronze)** |
| --- | --- | --- | --- |
| letreiro\_linha | StringType | Identificador público da linha (ex: "8000-10"). | l.c |
| codigo\_linha | LongType | Código interno/oficial da linha. | l.cl |
| prefixo\_onibus | LongType | Identificador único de cada ônibus. | vs.p |
| acessivel | BooleanType | true se o ônibus é adaptado para acessibilidade. | vs.a |
| timestamp\_captura | TimestampType | Data e hora exata (em UTC) em que a posição foi registrada. | vs.ta |
| latitude | DoubleType | A coordenada de latitude do ônibus. | vs.py |
| longitude | DoubleType | A coordenada de longitude do ônibus. | vs.px |
| ano | IntegerType | **(Coluna de Partição)** Ano extraído do timestamp\_captura. | Gerado pelo Spark |
| mes | IntegerType | **(Coluna de Partição)** Mês extraído do timestamp\_captura. | Gerado pelo Spark |
| dia | IntegerType | **(Coluna de Partição)** Dia extraído do timestamp\_captura. | Gerado pelo Spark |
| hora | IntegerType | **(Coluna de Partição)** Hora extraída do timestamp\_captura. | Gerado pelo Spark |