# Dicionário de Dados - Camada Gold

Este documento descreve a estrutura e os metadados das tabelas finais na Camada Gold, que é composta por duas partes: as tabelas de origem no **Lakehouse (MinIO)** e as tabelas de destino na **Camada de Servir (PostgreSQL)**.

A modelagem segue o padrão **Modelo Estrela (Star Schema)**.

## 1. Tabelas de Dimensão **(PostgreSQL)**

As tabelas de dimensão descrevem as entidades de negócio e residem no PostgreSQL para serem facilmente acessadas por todos os processos.

### Tabela: dim\_tempo

* **Tipo:** Dimensão
* **Atualização:** Estática. Populada uma única vez via script SQL.
* **Propósito:** Enriquecer os dados com atributos de tempo detalhados.
* **Chave Primária:** id\_tempo

| **Nome da Coluna** | **Tipo de Dado** | **Descrição** |
| --- | --- | --- |
| id\_tempo | SERIAL | Chave primária única para cada hora de cada dia. |
| data\_referencia | DATE | A data (ex: '2025-10-16'). |
| hora\_referencia | INTEGER | A hora do dia, em UTC (0-23). |
| ano | INTEGER | Ano da data de referência. |
| mes | INTEGER | Mês da data de referência. |
| dia | INTEGER | Dia da data de referência. |
| dia\_da\_semana | VARCHAR | O nome do dia da semana (ex: "Segunda-feira"). |
| fim\_de\_semana | BOOLEAN | true se for Sábado ou Domingo. |
| periodo\_do\_dia | VARCHAR | Classificação da hora (ex: "Manhã", "Tarde", "Noite", "Madrugada"). |

### Tabela: dim\_linha

* **Tipo:** Dimensão
* **Atualização:** Diariamente, pela DAG create\_dimensions do Airflow.
* **Propósito:** Fonte da verdade para os nomes e identificadores das linhas de ônibus.
* **Chave Primária:** id\_linha

| **Nome da Coluna** | **Tipo de Dado** | **Descrição** |
| --- | --- | --- |
| id\_linha | BIGINT | Chave primária única para cada linha. |
| letreiro\_linha | VARCHAR | O identificador público da linha (ex: "8000-10"). |
| nome\_linha | VARCHAR | O nome completo descritivo da linha. |

## 2. Tabelas de Fatos **no Lakehouse (MinIO)**

Estas são as tabelas principais e a **fonte única da verdade** para os dados agregados. Elas são armazenadas no formato **Delta Lake** para garantir transacionalidade (ACID) e versionamento.

### Tabela: fato\_operacao\_linhas\_hora

* **Localização:** s3a://gold/fato\_operacao\_linhas\_hora
* **Propósito:** Armazenar a contagem de ônibus únicos em operação por linha e por hora. Esta é a tabela principal da nossa Arquitetura Lambda.

| **Nome da Coluna** | **Tipo de Dado (Spark)** | **Descrição** |
| --- | --- | --- |
| id\_tempo | IntegerType | Chave estrangeira para dim\_tempo. |
| id\_linha | LongType | Chave estrangeira para dim\_linha. |
| quantidade\_onibus | LongType | Métrica com a contagem de ônibus. |

### Tabela: fato\_posicao\_onibus\_atual

* **Localização:** s3a://gold/fato\_posicao\_onibus\_atual
* **Propósito:** Manter o estado mais recente da posição de cada ônibus, servindo como a "memória" para os cálculos de KPI.

| **Nome da Coluna** | **Tipo de Dado (Spark)** | **Descrição** |
| --- | --- | --- |
| prefixo\_onibus | LongType | Identificador único do ônibus. |
| letreiro\_linha | StringType | Letreiro da linha associada. |
| codigo\_linha | LongType | Código da linha associada. |
| latitude | DoubleType | Coordenada de latitude. |
| longitude | DoubleType | Coordenada de longitude. |
| timestamp\_captura | TimestampType | Timestamp (UTC) da captura. |

### Tabela: fato\_velocidade\_linha

* **Localização:** s3a://gold/fato\_velocidade\_linha
* **Propósito:** Armazenar o KPI de velocidade operacional para cada linha na hora atual.

| **Nome da Coluna** | **Tipo de Dado (Spark)** | **Descrição** |
| --- | --- | --- |
| id\_tempo | IntegerType | Chave estrangeira para dim\_tempo. |
| id\_linha | LongType | Chave estrangeira para dim\_linha. |
| velocidade\_media\_kph | DoubleType | Métrica de velocidade em km/h. |
| updated\_at | TimestampType | Timestamp da última atualização do registro. |

### Tabela: fato\_onibus\_parados\_linha

* **Localização:** s3a://gold/fato\_onibus\_parados\_linha
* **Propósito:** Armazenar o KPI de contagem de ônibus parados/congestionados por linha na hora atual.

| **Nome da Coluna** | **Tipo de Dado (Spark)** | **Descrição** |
| --- | --- | --- |
| id\_tempo | IntegerType | Chave estrangeira para dim\_tempo. |
| id\_linha | LongType | Chave estrangeira para dim\_linha. |
| quantidade\_onibus\_parados | LongType | Métrica da contagem de ônibus parados. |
| updated\_at | TimestampType | Timestamp da última atualização do registro. |

## **3. Tabelas de Fatos na Camada de Servir (PostgreSQL)**

Estas tabelas são **cópias otimizadas** das tabelas do Lakehouse, carregadas ao final de cada pipeline. Elas são projetadas para consultas de baixa latência pela API e pelo Metabase. A estrutura das colunas é idêntica à das suas contrapartes no MinIO, mas com os tipos de dados do PostgreSQL (INTEGER, BIGINT, DOUBLE PRECISION, TIMESTAMP, etc.).