

---

**DW-BR**

*Release 0.1*

**Andre Pereira**

**18 jul. 2023**



<b>1</b>	<b>Contents</b>	<b>3</b>
1.1	Utilização do <i>Data Warehouse</i> . . . . .	3
1.2	Modelagem . . . . .	4
1.3	Tabelas . . . . .	4



**DW-BR** é uma armazém de dados públicos, que utiliza modelagem dimensional e ferramentas de *big data* com objetivo de oferecer acesso integrado, simples e intuitivo a conjuntos de dados públicos brasileiros.

Veja a Seção *Utilização do Data Warehouse* para mais informações sobre a instalação e utilização da ferramenta.

Detalhes sobre a modelagem, origem dos dados, metadados e outras informações técnicas estão disponíveis na Seção *Modelagem*.

---

**Nota:** Este projeto está em fase inicial de desenvolvimento.

---



## 1.1 Utilização do *Data Warehouse*

### 1.1.1 Instância local

Para definir uma instância local do DW-BR deve-se clonar o repositório e rodar os scripts de ETL da ferramenta:

```
$ git clone git@github.com:andrespp/dw-br.git
$ cd dw-br/
$ docker compose up -d
$ conda activate dwbr && ./get_ds.py && ./extract_ds.py && ./update-dw
```

### 1.1.2 Instância pública

```
>>> import dask.dataframe as dd
>>> df = dd.read_parquet('s3://bucket/my-parquet-data')
```

### 1.1.3 Ferramenta de Visualização Dash DW-BR

TBD

## 1.2 Modelagem

### 1.2.1 Conjuntos de dados

Os conjuntos de dados disponíveis no DW-BR são:

- Municípios Brasileiros
- CAGED

### 1.2.2 Dados

Dados originais (*Staging Tables*)

- stg\_caged

Tabelas Dimensão

- dim\_municipio

Telas Fato

## 1.3 Tabelas

### 1.3.1 stg\_caged

Origem dos dados

- CAGED

Atributos

Consultas Exemplo

```
SELECT
  COMPETENCIAMOV
, SUM(SALDOMOVIMENTACAO) AS SALDOMOV
FROM STG_CAGED
GROUP BY COMPETENCIAMOV
```

### 1.3.2 dim\_municipio

#### Origem dos dados - dim\_municipio

- Municipios Brasileiros

#### Atributos - dim\_municipio

#### Consultas Exemplo - dim\_municipio

```
SELECT
    UF, COUNT(COD_IBGE) AS QTD_MUNICIPIOS
FROM DIM_MUNICIPIO
GROUP BY UF
ORDER BY COUNT(COD_IBGE) DESC
```