

# Documentação do Data Warehouse - Portal de Estatística da Madeira (DREM)

## Visão geral:

O Data Warehouse do Portal de Estatística da DREM foi projetado para gerir e disponibilizar dados estatísticos da Região Autónoma da Madeira de forma eficiente e escalável. A estrutura suporta análises multidimensionais complexas, integração com IA, e apresentação de dados em múltiplos idiomas.

## Estatísticas Gerais:

- Temas: 4
- Sub-temas: 18
- Áreas: 93

TEMAS	SUB_TEMAS	AREAS
1- Economia	Id: 1- 12 (total: 12)	Id: 1-55 (total: 55)
2- Social	Id: 13-17 (total: 5)	Id: 56-75 (total: 20)
3- A Região	Id: 18, 19 (total: 2)	Id: 76-86 (total: 11)
4- Multitemas	N/A	Id: 87-93 (total: 7)
<b>TOTAL: 4 (Temas)</b>	<b>18 (Sub-temas)</b>	<b>93 (Áreas)</b>

## Objetivos Principais:

- **Análise de Dados:** Permitir análises detalhadas através de estruturas OLAP.
- **Flexibilidade e Crescimento:** Estrutura modular para suportar novos temas e indicadores.
- **Suporte Multilíngua:** Dados disponíveis em português e inglês.
- **Integração com IA:** Estrutura amigável para modelos de IA e chatbots.

## 1- Arquitetura do Datawarehouse

O DW utiliza um modelo híbrido que combina:

- **Esquema Estrela:** Para análises dimensionais eficientes
- **Esquema Floco de Neve:** Para hierarquias complexas
- **Bridge Tables:** Para dimensões dinâmicas

## 2- Estrutura das tabelas:

Coluna	Tipo	Obrigatório	Descrição
Nome da coluna	Tipo de dados	Sim / Não	Descrição em português.

**Tabela 1:** temas

- Tabela que armazena os temas principais dos dados.

Coluna	Tipo	Obrigatório	Descrição
id_tema	UUID	Sim	Identificador único do tema.
codigo	VARCHAR(50)	Sim	Código único do tema.
nome_pt	VARCHAR(255)	Sim	Nome do tema em português.
nome_en	VARCHAR(255)	Não	Nome do tema em inglês.
descricao_pt	TEXT	Não	Descrição do tema em português.
descricao_en	TEXT	Não	Descrição do tema em inglês.
icone	VARCHAR(100)	Não	Ícone associado ao tema.
cor_tema	VARCHAR(7)	Não	Cor associada ao tema (em formato hexadecimal).
metadata	JSONB	Não	Informações adicionais sobre o tema.
ordem	INTEGER	Não	Ordem de apresentação do tema.
ativo	BOOLEAN	Não	Indica se o tema está ativo ou não.
slug_pt	VARCHAR(255)	Automático	URL amigável do nome do tema em português.
slug_en	VARCHAR(255)	Automático	URL amigável do nome do tema em inglês.
created_at	TIMESTAMP WITH TIME ZONE	Sim	Data de criação do tema.
updated_at	TIMESTAMP WITH TIME ZONE	Sim	Data de última atualização do tema

**Explicação:** A tabela temas contém informações sobre os temas principais utilizados para organizar os dados. Os temas atuais são “Economia”, “Social”, “Regiao”, “Multitemas” . A tabela armazena metadados como ícones e cores para facilitar a visualização.

**Lista atual dos Temas:** 1- Economia; 2- Social; 3- Região; 4- Multitemas

## Tabelas 2: sub\_temas

- Tabela que armazena sub-temas organizados hierarquicamente sob cada tema.

Coluna	Tipo	Obrigatório	Descrição
id_sub_tema	UUID	Sim	Identificador único do subtema.
id_tema	UUID	Sim	Identificador do tema relacionado.
codigo	VARCHAR(50)	Sim	Código único do subtema.
nome_pt	VARCHAR(255)	Sim	Nome do subtema em português.
nome_en	VARCHAR(255)	Não	Nome do subtema em inglês.
descricao_pt	TEXT	Não	Descrição do subtema em português.
descricao_en	TEXT	Não	Descrição do subtema em inglês.
path	ltree	Sim	Caminho hierárquico do subtema.
metadata	JSONB	Não	Informações adicionais sobre o subtema.
ordem	INTEGER	Não	Ordem de apresentação do subtema.
ativo	BOOLEAN	Não	Indica se o subtema está ativo ou não.
slug_pt	VARCHAR(255)	Automático	URL amigável do nome do subtema em português.
slug_en	VARCHAR(255)	Automático	URL amigável do nome do subtema em inglês.
created_at	TIMESTAMP WITH TIME ZONE	Sim	Data de criação do subtema.
updated_at	TIMESTAMP WITH TIME ZONE	Sim	Data de última atualização do subtema.

**Explicação:** A tabela sub\_temas organiza temas em categorias menores. Por exemplo, dentro de "Economia", temos o sub-tema "Turismo".

**Lista atual dos sub-temas:** 1- Administração Pública; 2- Agricultura, floresta e pesca

3-Comércio; 4- Conjuntura; 5- Construção e Habitação; 6- Contas Económicas; 7- Empresas

8- Indústria e energia; 9- Inovação e conhecimento; 10- Sector monetário e financeiro

11- Transportes e Comunicações, 12- Turismo, 13- Condições de vida, 14- Educação e formação

15- Mercado de Trabalho, 16- População, 17- Saúde, 18- Justiça

**Tabela 3:** areas

- Tabela que detalha áreas específicas dentro de sub-temas.

Coluna	Tipo	Obrigatório	Descrição
id_area	UUID	Sim	Identificador único da área.
id_sub_tema	UUID	Sim	Identificador do subtema relacionado.
codigo	VARCHAR(50)	Sim	Código único da área.
nome_pt	VARCHAR(255)	Sim	Nome da área em português.
nome_en	VARCHAR(255)	Não	Nome da área em inglês.
descricao_pt	TEXT	Não	Descrição da área em português.
descricao_en	TEXT	Não	Descrição da área em inglês.
icone	VARCHAR(100)	Não	Ícone associado à área.
path	ltree	Não	Caminho hierárquico da área.
metadata	JSONB	Não	Informações adicionais sobre a área.
ordem	INTEGER	Não	Ordem de apresentação da área.
ativo	BOOLEAN	Não	Indica se a área está ativa ou não.
slug_pt	VARCHAR(255)	Automático	URL amigável do nome da área em português.
slug_en	VARCHAR(255)	Automático	URL amigável do nome da área em inglês.
created_at	TIMESTAMP WITH TIME ZONE	Sim	Data de criação da área.
updated_at	TIMESTAMP WITH TIME ZONE	Sim	Data de última atualização da área.

**Explicação:** Esta tabela guarda áreas específicas dentro de cada subtema. Por exemplo, dentro do subtema "Turismo" podemos ter as áreas: "Oferta Turística" e "Gastos Turísticos Internacionais".

**Lista atual dos sub-temas:** 1- Administração Local; 2- Emprego Público; 3- Procedimentos dos Défices Excessivos; 4- Receita e Despesa Pública; 5- Receitas Fiscais; 6- Comercialização de banana; 7- Estatísticas anuais; 8- Floresta; 9 – Floresta; 10 – Floricultura; 11- Produção animal e pesca; 12- Recenseamentos e inquéritos agrícolas estruturais; 13- Comercialização de produtos tradicionais da Madeira; 14- Comércio Internacional; 15- Comércio interno (inclui as UCDR); 16- COVID-19 / Guerra na Ucrânia; 17- Inquéritos Qualitativos de Conjuntura; 18- Indicadores mensais de conjuntura; 19- Crédito à habitação; 20- Estatísticas anuais ; 21- Habitação Social ; 22- Indicadores das empresas da construção; 23- Licenciamento e conclusão de obras ; 24- Operações sobre imóveis ; 25- Preços na habitação; 26- Venda de alojamentos familiares ; 27- Venda de cimento; 28- Contas Económicas Regionais da Agricultura; 29- Contas Regionais; 30- Conta Satélite do Mar; 31- Conta Satélite do Turismo; 32- Investimento Direto Estrangeiro; 33- Matriz input-output; 34- Contas integradas das empresas; 35- Custos de contexto; 36- Necessidades de qualificações; 37- Práticas de gestão; 38- Serviços prestados às empresas; 39- Sociedades constituídas e dissolvidas;

40- Balanço energético; 41- Combustíveis; 42- Consumo de energia no sector doméstico; 43- Energia elétrica; 44- Indústria ; 45- Inovação; 46- Investigação e desenvolvimento; 47- Sociedade de informação; 48- Preços no Consumidor; 49- Banca e seguros-; 50- Empréstimos; 51- Rede SIBS; 52- Comunicações; 53- Transportes; 54- Oferta turística; 55- Gastos turísticos internacionais; 56- Pobreza e desigualdade; 57- Rendimento e despesa; 58- Cultura e Desporto; 59- Educação e Formação de Adultos; 60- Estatísticas anuais; 61- Acidentes de trabalho; 62- Custo de trabalho; 63- Emprego, desemprego e inatividade; 64- Quadros de Pessoal; 65- Remunerações; 66- Trabalho voluntário; 67- Censos; 68- Demografia; 69- Fecundidade; 70- Projeções da População Residente; 71- Tábuas de mortalidade; 72- Proteção Social; 73- Estatísticas anuais; 74- Inquérito Nacional de Saúde; 75- Ocorrências pré-hospitalares; 76- Ambiente; 77- Dados meteorológicos; 78- Geografia física e humana; 79- Índice sintético de desenvolvimento territorial; 80- Estatísticas Anuais; 81- Falências e Insolvências; 82- Vitimação; 83- Mar; 84- Participação política; 85- Poder de compra concelho; 86- Retrato territorial; 87- Anuário estatístico; 88- Atlas Estatístico; 89- Barómetro das RUP; 90- Boletim trimestral; 91- Carta de Equipamentos e Serviços de Apoio à População; 92- Indicadores do Portugal 2020; 93- Madeira em Números.

**Tabela 4:** indicadores

- Tabela de indicadores, que contém métricas e dados estatísticos relevantes para análise.

Coluna	Tipo	Obrigatório	Descrição
id_indicador	UUID	Sim	Identificador único do indicador.
id_area	UUID	Sim	Referência à área correspondente.
codigo	VARCHAR(50)	Sim	Código único do indicador.
nome_pt	VARCHAR(255)	Sim	Nome do indicador em português.
nome_en	VARCHAR(255)	Não	Nome do indicador em inglês (opcional).
descricao_pt	TEXT	Não	Descrição do indicador em português (opcional).
descricao_en	TEXT	Não	Descrição do indicador em inglês (opcional).
unidade	VARCHAR(50)	Sim	Unidade de medida do indicador, não pode estar vazia.
tipo_valor	valor_tipo	Sim	Tipo do valor do indicador (ex: número, percentagem, moeda, texto).
estado	estado_indicador	Sim	Estado do indicador (preliminar, provisório, definitivo)
periodicidade	periodicidade_tipo	Sim	Frequência de atualização do indicador (ex: diária, mensal, trimestral).
fonte	VARCHAR(255)	Sim	Fonte de dados do indicador.
metodologia_pt	TEXT	Não	Metodologia em português utilizada para calcular o indicador (opcional).
metodologia_en	TEXT	Não	Metodologia em inglês utilizada para calcular o indicador (opcional).
formula	TEXT	Não	Fórmula utilizada para calcular o indicador (opcional).
dimensoes_aplicaveis	UUID[]	Sim	Lista de dimensões aplicáveis ao indicador (não pode ser nula).
metadata	JSONB	Não	Metadados adicionais do indicador (opcional).
configuracao_visualizacao	JSONB	Não	Configurações de visualização do indicador (opcional).
configuracao_olap	JSONB	Não	Configurações OLAP para o indicador (opcional).

<b>palavras_chave_pt</b>	TEXT[]	Não	Palavras-chave em português associadas ao indicador (opcional).
<b>palavras_chave_en</b>	TEXT[]	Não	Palavras-chave em inglês associadas ao indicador (opcional).
<b>valor_minimo</b>	NUMERIC	Não	Valor mínimo do indicador (opcional).
<b>valor_maximo</b>	NUMERIC	Não	Valor máximo do indicador (opcional).
<b>precision_scale</b>	INTEGER	Não	Precisão do valor (0 a 6), padrão 2.
<b>frequencia_atualizacao</b>	INTERVAL	Não	Frequência de atualização do indicador (opcional).
<b>ultima_atualizacao</b>	TIMESTAMP WITH TIME ZONE	Não	Data e hora da última atualização do indicador (opcional).
<b>proxima_atualizacao</b>	TIMESTAMP WITH TIME ZONE	Não	Data e hora da próxima atualização do indicador (opcional).
<b>versao</b>	INTEGER	Não	Versão do indicador, padrão 1.
<b>ativo</b>	BOOLEAN	Não	Indica se o indicador está ativo ou não, padrão verdadeiro.
<b>created_at</b>	TIMESTAMP WITH TIME ZONE	Não	Data e hora de criação do registro.
<b>updated_at</b>	TIMESTAMP WITH TIME ZONE	Data da última atualização do registro.	updated_at

**Explicação:** Esta tabela armazena informações sobre diferentes indicadores que podem ser usados para monitorizar e avaliar variáveis específicas em diversas áreas, como saúde, educação, economia e meio ambiente. Cada indicador possui um código único, um nome (em português e, opcionalmente, em inglês), uma descrição, e detalhes sobre a unidade de medida. Além disso, inclui informações sobre a periodicidade de atualização, a fonte dos dados, e a metodologia usada para calcular o indicador. Os indicadores também estão associados a dimensões que ajudam a contextualizar os dados, como o tempo e a geografia, permitindo análises mais detalhadas e significativas.

**Tabela 5:** dados\_indicadores

- Tabela de dados de indicadores, que registra os valores associados a cada indicador em diferentes períodos e contextos.

Coluna	Tipo	Obrigatório	Descrição
id_dado	UUID	Sim	Identificador único do dado.
id_indicador	UUID	Sim	Referência ao indicador correspondente.
valor	NUMERIC	Não	Valor do indicador; deve ser maior ou igual a zero, se não for nulo.
valor_anterior	NUMERIC	Não	Valor do indicador no período anterior (opcional).
variacao_percentual	NUMERIC	Não	Variação percentual entre o valor atual e o anterior, calculada automaticamente.
intervalo_confianca_min	NUMERIC	Não	Limite inferior do intervalo de confiança (opcional).
intervalo_confianca_max	NUMERIC	Não	Limite superior do intervalo de confiança (opcional).
flags	TEXT[]	Não	Indicadores adicionais para análise do dado (opcional).
fonte_especifica	VARCHAR(255)	Não	Fonte de dados específica do indicador
metadata	JSONB	Não	Metadados adicionais do dado (opcional).
created_at	TIMESTAMP WITH TIME ZONE	Não	Data e hora de criação do registro.
updated_at	TIMESTAMP WITH TIME ZONE	Não	Data e hora da última atualização do registro.

**Explicação:** Esta tabela registra os valores numéricos dos indicadores, incluindo referências a dimensões para análise contextualizada. Os dados podem conter valores atuais, anteriores, variações percentuais e intervalos de confiança. É essencial para a análise estatística, fornecendo dados brutos para relatórios e insights sobre tendências dos indicadores ao longo do tempo, organizados para facilitar comparações e análises.



**Tabela 6:** dimensoes

- Tabela de dimensões, que categoriza os dados para análise multidimensional.

Coluna	Tipo	Obrigatório	Descrição
id_dimensao	UUID	Sim	Identificador único da dimensão.
id_parent	UUID	Não	Identificador da dimensão pai (se houver).
tipo	dimensao_tipo	Sim	Tipo da dimensão (geográfica, temporal, etc.).
codigo	VARCHAR(50)	Sim	Código único da dimensão.
nome_pt	VARCHAR(255)	Sim	Nome da dimensão em português.
nome_en	VARCHAR(255)	Não	Nome da dimensão em inglês.
valor	VARCHAR(255)	Sim	Valor específico da dimensão.
hierarquia	ltree	Sim	Hierarquia da dimensão para análise.
nivel	INTEGER	Sim	Nível hierárquico da dimensão.
ordem	INTEGER	Não	Ordem de apresentação da dimensão.
data_inicio	DATE	Não	Data de início da validade da dimensão
data_fim	DATE	Não	Data de fim da validade da dimensão
metadata	JSONB	Não	Informações adicionais sobre a dimensão
ativo	BOOLEAN	Sim	Indica se a dimensão está ativa ou não.
created_at	TIMESTAMP WITH TIME ZONE	Sim	Data de criação da dimensão.
updated_at	TIMESTAMP WITH TIME ZONE	Sim	Data de última atualização da dimensão.

**Explicação:** Esta tabela guarda as diferentes formas de classificar e organizar os dados. Por exemplo, podemos ter dimensões de tempo (anos, meses), geografia (países, regiões), demografia (faixas etárias, género), etc.

**Tabela 7:** indicador\_dimensoes\_bridge

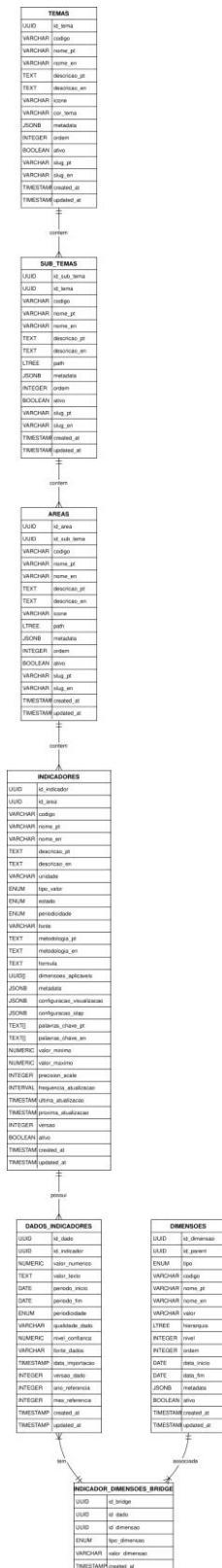
- Tabela de “ponte” para lidar com dimensões dinâmicas

Coluna	Tipo	Obrigatório	Descrição
id_brige	UUID	Sim	Identificador único da dimensão.
id_dado	UUID	Sim	Identificador da dimensão pai (se houver).
id_dimensao	UUID	Sim	Tipo da dimensão (geográfica, temporal, etc.).
tipo_dimensao	ENUM(dimensao_tipo)	Sim	Código único da dimensão.
valor_dimensao	VARCHAR(255)	Sim	Nome da dimensão em português.
created_at	TIMESTAMP WITH TIME ZONE	Sim	Data de criação da dimensão.
updated_at	TIMESTAMP WITH TIME ZONE	Sim	Data de última atualização da dimensão.

**Explicação:**

- Em vez de ter colunas fixas, cada relação de dimensão é armazenada como uma linha separada na tabela ponte
- Um facto (dado) pode ter qualquer número de relações de dimensão
- Podem ser adicionados novos tipos de dimensão sem alterar o esquema da base de dados
- Indicadores diferentes podem utilizar combinações diferentes de dimensões

## 2- Diagrama ER do Datawarehouse



### 3- Análise da Estrutura

#### 1. Organização Intuitiva

- **Funciona como uma "árvore" organizada:**

Temas → Sub-temas → Áreas → Indicadores

- **Exemplo:** Economia → Turismo → Oferta turística → Total de Hóspedes (nº)

#### 2. Preparada para Crescimento

- **Estrutura modular:** Novos temas e indicadores podem ser adicionados sem impacto nas tabelas existentes.
- **Escalabilidade:** A base é projetada para suportar milhões de registros com bom desempenho.

#### 3. Integração com IA (Chatbot)

- **Organização amigável:** Organização de forma que modelos de IA possam:
  - Encontrar dados de maneira eficiente.
  - Compreender a relação entre diferentes indicadores.
  - Fornecer contexto nas respostas automaticamente.

#### 4. Multilíngua

- **Suporte a múltiplos idiomas:** Dados disponíveis em Português e Inglês, com facilidade para adicionar outros idiomas no futuro.

#### 5. Visualização Flexível

- **Pronta para gráficos e tabelas:** Configurações de visualização são guardadas junto com os dados, permitindo integração direta com ferramentas como Highcharts e WebDataRocks.

#### 6. Qualidade dos Dados

- **Rastreamento de origem e validação automática:** Controle da qualidade dos dados através de validações e rastreamento de alterações.

#### 7. Desempenho

- **Partição e caching:** Melhoria na performance para consultas rápidas, especialmente em grandes volumes de dados.

## 4- Benefícios para o Projeto do Chatbot com LLM

### 1. Contexto Rico

- **Compreensão hierárquica:** O chatbot pode entender toda a estrutura e responder com precisão.
  - **Exemplo:** Ao ser perguntado sobre "turismo", o chatbot consegue relacionar todos os indicadores do tema.

### 2. Respostas Precisas

- **Informações detalhadas e validadas:** Capaz de fornecer dados exatos, incluindo a fonte e metodologia.

### 3. Múltiplos Formatos

- **Adaptabilidade:** Dados podem ser apresentados em diferentes formatos (tabelas, gráficos, etc.).

### 4. Atualizações Fáceis

- **Manutenção simplificada:** Novos dados são integrados automaticamente, facilitando a atualização contínua.