



**Politécnico
Castelo Branco**

Escola Superior
de Tecnologia

Implementação de um Data Warehouse Para Análise de Consumo Elétrico

Francisco Gonçalves Nº20221849

Jorge Nunes Nº20221294

Docente

Filipe Fidalgo

Janeiro 2026

Resumo

Este trabalho tem como objetivo a implementação de um Data Warehouse para análise do consumo de energia elétrica, aplicando os conceitos e metodologias de Data Warehousing abordados na unidade curricular. O sistema desenvolvido pretende suportar a análise multidimensional dos dados e apoiar a tomada de decisão, através da exploração histórica da informação.

A solução baseia-se num modelo multidimensional em esquema de estrela, composto por uma tabela de factos que armazena o consumo energético e por dimensões que permitem a análise segundo diferentes perspetivas, nomeadamente tempo, localização geográfica e nível de tensão.

Palavras chave

Data Warehouse, modelagem Multidimensional, Esquema em Estrela, consumo de energia.

Abstract

This work aims to implement a Data Warehouse for the analysis of electric energy consumption, applying the concepts and methodologies of Data Warehousing covered in the course unit. The developed system is intended to support multidimensional data analysis and decision-making, through the historical exploration of information.

The solution is based on a multidimensional star schema, composed of a fact table that stores energy consumption and several dimensions that allow analysis from different perspectives, namely time, geographical location, and voltage level.

Keywords

Data Warehouse, Multidimensional Modeling, Star Schema, Energy consumption.

V

Índice geral

1. Introdução.....	1
2. Caracterização da Organização e do Problema	2
3. Definição da Arquitetura do Sistema	3
4. Objetivos do <i>Data Warehouse</i>	4
4.1. Questões de Negócio	4
5. Modelação Multidimensional	5
5.1. Abordagem Utilizada	5
5.2. Diagrama do Modelo Multidimensional	5
5.3. Implementação da <i>Data Warehouse</i>	6
6. Processo de ETL.....	8
6.1. Fonte de Dados	8
6.2. Extração	8
6.3. Transformação e Carregamento.....	8
7. Exploração Analítica em SQL.....	10
7.1. Índices.....	10
7.2. Common Table Expressions (CTE)	11
7.3. ROLLUP	11
7.4. CUBE	12
7.5. GROUPING SETS.....	12
8. Análise em Excel (PowerPivot).....	14
8.1. Consumo por Distrito e Ano	14
8.2. Consumo total por Distrito e Mês	15
8.3. Consumo por Distrito e Nível de Tensão.....	16
8.4. Consumo por Ano e Nível de Tensão	17
9. Visualização em PowerBI.....	18
9.1. Consumo por Distrito e Ano	18
9.2. Consumo total por Distrito e Mês	19
9.3. Consumo por Distrito e Nível de Tensão.....	20
9.4. Consumo por Ano e Nível de Tensão	21
10. Relatórios de Apoio à Decisão?.....	22
11. Conclusão.....	23

12. Referencias.....	24
13. Anexo.....	25
13.1. Drops.....	25
13.2. Criar <i>Data Warehouse</i>	28
13.3. Transformação e Carregamento	33
13.4. <i>Queries SQL</i>	37

Índice de figuras

Figura 1 - Arquitetura do Sistema.....	3
Figura 2 – Modelo Multidimensional.....	6
Figura 3 – Estrutura física da Data Warehouse com separação por filegroups	7
Figura 4 - PowerPivot Consumo por Distrito e Ano	14
Figura 5 - PowerPivot Consumo total por Distrito e Mês	15
Figura 6 - PowerPivot Consumo total por Distrito e Nível de Tensão	16
Figura 7 - PowerPivot Consumo total por Ano e Nível de Tensão.....	17
Figura 8 - PowerBI Consumo por Distrito e Ano.....	18
Figura 9 - PowerBI Consumo total por Distrito e Mês	19
Figura 10 - PowerBI Consumo total por Distrito e Nível de Tensão	20
Figura 11 - PowerBI Consumo total por Ano e Nível de Tensão	21

1. Introdução

Nos últimos anos, o crescimento exponencial do volume de dados gerado pelas organizações tornou indispensável a adoção de soluções capazes de transformar dados em informação relevante para apoio à decisão. Neste contexto, o Data Warehousing assume um papel fundamental, ao permitir a integração e análise estruturada de dados históricos provenientes de múltiplas fontes.

A análise histórica constitui uma das principais mais-valias de um Data Warehouse, possibilitando a identificação de padrões e tendências ao longo do tempo, através de consultas analíticas complexas. Esta abordagem revela-se relevante no setor da energia, onde o estudo do consumo elétrico por período temporal, localização ou características técnicas contribui para uma gestão mais eficiente dos recursos.

O presente trabalho enquadra-se na unidade curricular de Data Warehousing, do Mestrado em Engenharia Informática, e tem como objetivo a implementação de uma Data Warehouse com base num dataset da empresa E-Redes [1], relativo ao consumo de eletricidade [2], suportando a modelação multidimensional e os processos de ETL.

2. Caracterização da Organização e do Problema

A E-Redes foi utilizada neste trabalho como organização de referência, servindo de exemplo realista para a conceção e implementação de uma *Data Warehouse* no domínio da energia elétrica. A escolha desta entidade prende-se com a relevância do setor energético e com a disponibilidade pública de dados relacionados com o consumo de eletricidade, permitindo enquadrar o trabalho num contexto próximo de um cenário organizacional real.

Para a realização do trabalho, foi obtido um *dataset* em formato Excel a partir do website oficial da E-Redes [1], contendo dados históricos de consumo de energia elétrica [2]. Este conjunto de dados inclui, entre outros, os seguintes atributos: Ano, Mês, Data, Distrito, Concelho, Freguesia, Nível de Tensão e Energia Ativa (kWh), bem como códigos administrativos e campos auxiliares. A informação disponibilizada permite analisar o consumo elétrico segundo diferentes perspetivas temporais, geográficas e técnicas, constituindo uma base adequada para a construção de um modelo multidimensional orientado à análise analítica.

Apesar de o formato Excel permitir uma análise inicial dos dados, este apresenta limitações significativas quando se pretende efetuar análises históricas mais complexas e sobre volumes elevados de informação. Neste contexto, surge a motivação para a criação de uma *Data Warehouse*, cujo objetivo principal é centralizar, estruturar e otimizar a análise dos dados de consumo elétrico. A implementação da DW permite separar o processamento analítico de abordagens mais simples ou operacionais, suportando consultas complexas de forma eficiente através de um modelo multidimensional.

3. Definição da Arquitetura do Sistema

A arquitetura da *Data Warehouse* desenvolvida neste trabalho segue uma abordagem clássica de *Data Warehousing*, apresentada na **Figura 1**, organizada em três camadas principais: fontes de dados, área de *staging* e *Data Warehouse*. Como fonte de dados foi utilizado um ficheiro em formato Excel, disponibilizado pela E-Redes, contendo dados históricos relativos ao consumo de energia elétrica. Estes dados são inicialmente carregados para uma área de *staging*, onde são armazenados de forma próxima à origem. A partir da área de *staging*, os dados são transformados e carregados para a *Data Warehouse*, modelada segundo um esquema em estrela.

Esta estrutura permite uma separação clara entre a tabela de factos, onde são armazenadas as métricas de consumo, e as várias dimensões que contextualizam a análise, nomeadamente tempo, localização geográfica e nível de tensão. Esta arquitetura foi concebida para suportar análises históricas eficientes e adequadas ao apoio à decisão.

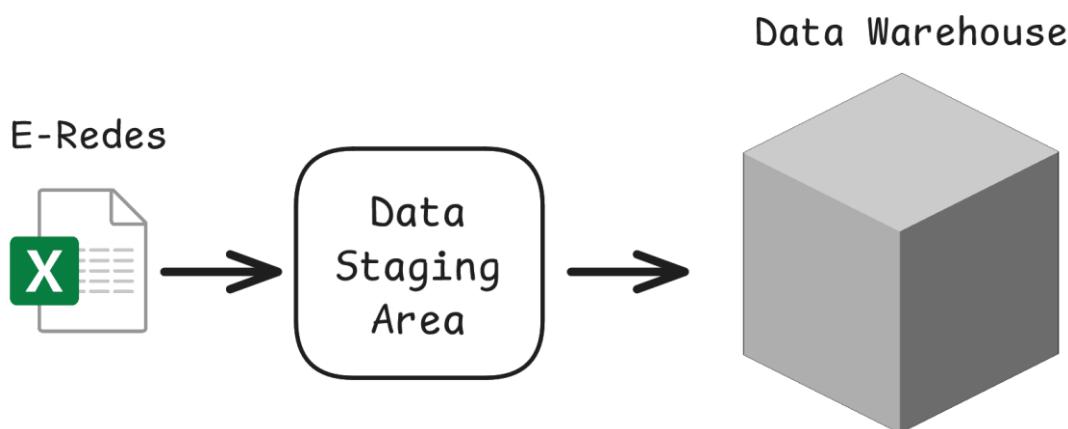


Figura 1 - Arquitetura do Sistema

4. Objetivos do *Data Warehouse*

O principal objetivo da *Data Warehouse* desenvolvida neste trabalho é suportar a análise histórica do consumo de energia elétrica, com base em dados provenientes da E-Redes, permitindo a exploração da informação segundo diferentes perspetivas temporais, geográficas e técnicas. A *Data Warehouse* foi concebida para centralizar e estruturar os dados de consumo, garantindo consistência, desempenho e flexibilidade na execução de consultas analíticas.

Adicionalmente, a *Data Warehouse* pretende facilitar a realização de análises agregadas e comparativas, permitindo a identificação de padrões de consumo ao longo do tempo e a geração de informação relevante para apoio à decisão. A utilização de um modelo multidimensional em estrela, aliada a mecanismos de otimização como índices e particionamento, visa assegurar tempos de resposta adequados mesmo na execução de consultas analíticas mais complexas.

4.1. Questões de Negócio

Com base nos objetivos definidos e na estrutura da *Data Warehouse* implementada, foram identificadas várias questões de negócio que a solução pretende responder, refletindo necessidades típicas de análise no setor da energia:

- **Qual é o consumo de energia elétrica por distrito e por ano?**
Permite analisar a evolução do consumo energético ao longo do tempo em diferentes regiões, identificando tendências e variações anuais.
- **Qual é o consumo total por distrito e por mês, incluindo subtotais?**
Possibilita uma análise mais detalhada do consumo mensal, bem como a obtenção de subtotais que facilitam comparações entre regiões e períodos temporais.
- **Qual é o consumo de energia por distrito e por nível de tensão?**
Permite avaliar o impacto do consumo consoante o nível de tensão da rede, contribuindo para uma melhor compreensão da distribuição do consumo energético.
- **Qual é o consumo de energia por ano e por nível de tensão?**
Facilita a análise comparativa entre diferentes anos e níveis de tensão, apoiando a identificação de padrões e alterações no comportamento do consumo ao longo do tempo.

5. Modelação Multidimensional

5.1. Abordagem Utilizada

A modelação multidimensional da *Data Warehouse* foi realizada com base numa abordagem em esquema estrela, amplamente utilizada em sistemas de apoio à decisão devido à sua simplicidade, clareza e eficiência na execução de consultas analíticas. Esta abordagem permite separar de forma clara os dados quantitativos, armazenados na tabela de factos, da informação descritiva, organizada em tabelas de dimensões.

No contexto deste trabalho, a modelação foi orientada pelas questões de negócio previamente definidas, garantindo que a estrutura da *Data Warehouse* suporta análises temporais, geográficas e tensão elétrica.

A tabela de factos armazena a métrica principal do sistema, o consumo de energia elétrica, enquanto as dimensões fornecem o enquadramento necessário para análise dos dados temporais, geográficos e níveis de tensão utilizados.

5.2. Diagrama do Modelo Multidimensional

O modelo multidimensional implementado é constituído por uma tabela de factos central, *fact_consumo_eletrico*, ligada a três dimensões principais: Tempo, Localização e Voltagem, conforme ilustrado na **Figura 2**.

A dimensão Tempo (*dim_tempo*) permite a análise do consumo de energia ao longo do tempo, disponibilizando diferentes níveis de granularidade, ano, mês, trimestre e semestre. Esta dimensão é fundamental para a realização de análises históricas e comparativas.

A dimensão Localização (*dim_localizacao*) representa a hierarquia geográfica dos dados, incluindo distrito, concelho e freguesia, possibilitando a análise do consumo por diferentes níveis territoriais. Esta estrutura suporta consultas que permitem comparar padrões de consumo entre regiões.

A dimensão Voltagem (*dim_voltagem*) caracteriza o nível de tensão da rede elétrica associado ao consumo, permitindo análises técnicas relacionadas com o tipo de fornecimento de energia.

A tabela de factos (*fact_consumo_eletrico*) armazena o valor do consumo de energia elétrica, expresso em MWh, e contém chaves estrangeiras que estabelecem a ligação às respetivas dimensões.

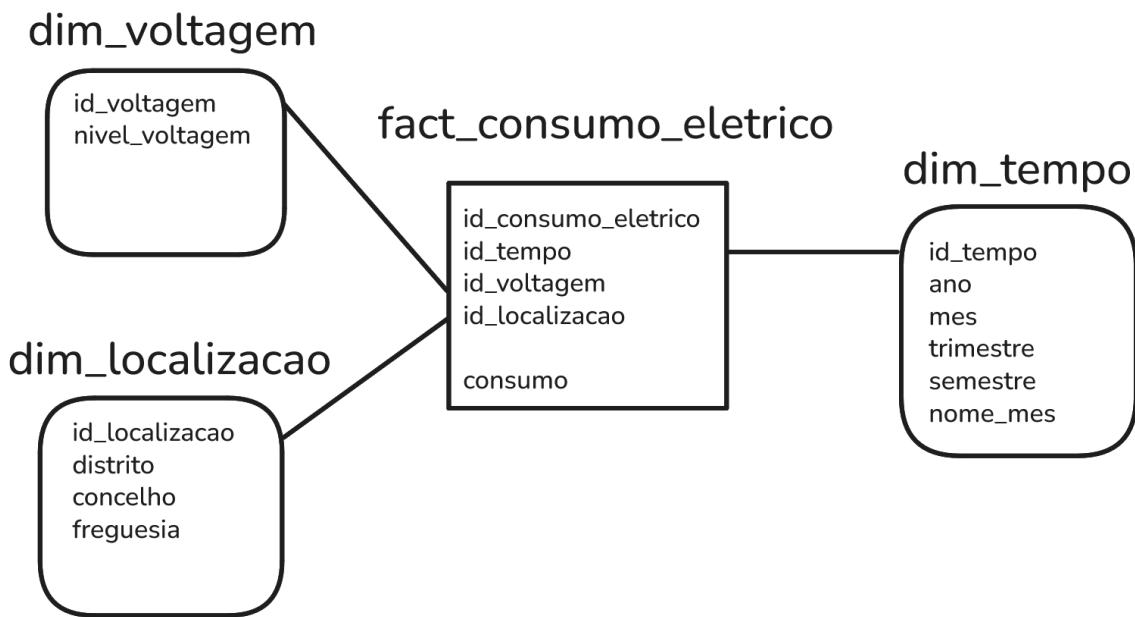


Figura 2 – Modelo Multidimensional

5.3. Implementação da *Data Warehouse*

A implementação da *Data Warehouse* foi realizada no Microsoft SQL Server, tendo sido consideradas não só a criação das estruturas lógicas do modelo multidimensional, mas também decisões ao nível da organização física dos dados, com o objetivo de melhorar o desempenho e a escalabilidade.

Foram definidos *filegroups* distintos para separar os dados dimensionais, as chaves primárias e a tabela de factos. As dimensões foram alocadas a um *filegroup* próprio, enquanto a tabela de factos foi distribuída por dois *filegroups* distintos, permitindo uma melhor organização e gestão do armazenamento.

A tabela de factos foi particionada com base na dimensão Tempo, utilizando o atributo id_tempo. Para esse efeito, os registos anteriores a 2022 foram associados a identificadores temporais negativos, enquanto os registos posteriores foram associados a identificadores positivos. Este critério permitiu separar os dados históricos dos dados mais recentes, distribuindo-os por diferentes partições e *filegroups*.

A **Figura 3** apresenta a distribuição dos dados descrita pelos *filegroups*.

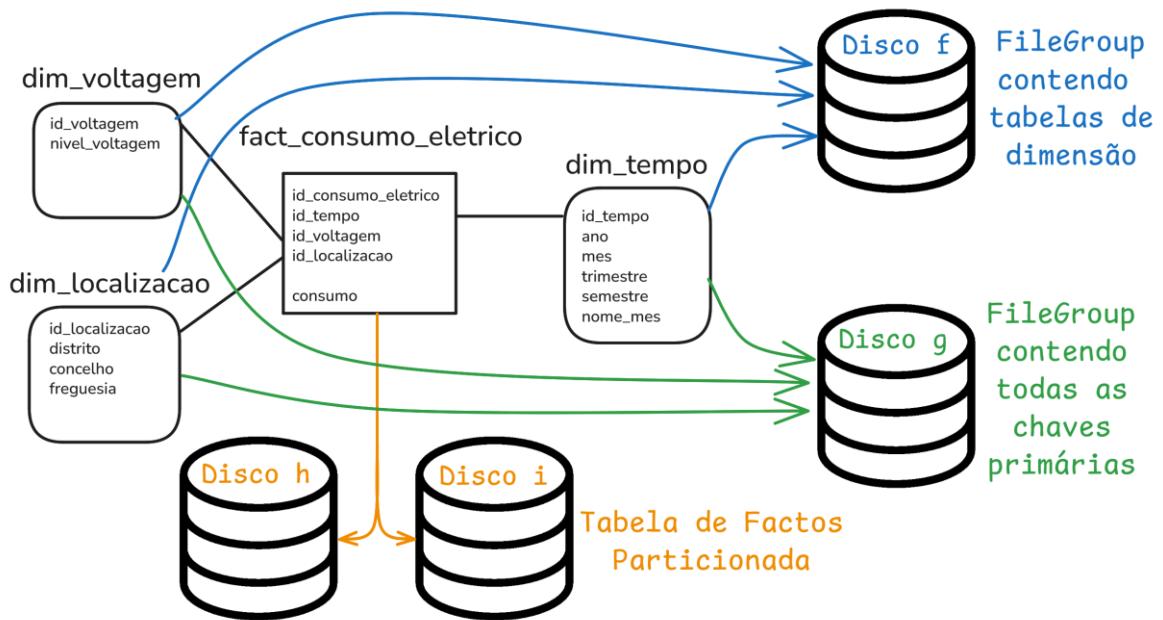


Figura 3 – Estrutura física da Data Warehouse com separação por filegroups

6. Processo de ETL

6.1. Fonte de Dados

A fonte de dados utilizada no processo de ETL consiste num *dataset* em formato Excel, disponibilizado publicamente no website da E-Redes [1], contendo informação histórica relativa ao consumo de energia elétrica [2]. O ficheiro apresenta os dados organizados por registo de consumo, integrando atributos temporais, geográficos e técnicos, permitindo uma análise detalhada do consumo elétrico.

Cada registo do *dataset* disponibilizado inclui um conjunto de atributos de natureza temporal, geográfica e técnica, nomeadamente: Ano, Mês, Data, Distrito, Concelho, Freguesia, Nível de Tensão, Energia Ativa (kWh), CodDistrito, CodDistritoConcelho, CodDistritoConcelhoFreguesia e o campo auxiliar mes_int, correspondente à representação numérica do mês.

6.2. Extração

No âmbito do processo de ETL, a fase de extração consistiu na obtenção dos dados a partir do ficheiro Excel disponibilizado pela E-Redes, os quais foram carregados para uma base de dados de *staging*, criada especificamente para este fim. Esta área de *staging* funciona como uma camada intermédia entre a fonte de dados e a *Data Warehouse*, permitindo armazenar os dados de forma próxima ao formato original e facilitando a sua validação e preparação para as fases seguintes do processo.

6.3. Transformação e Carregamento

A fase de transformação e carregamento teve como objetivo preparar os dados armazenados na área de *staging* para integração na *Data Warehouse*, de acordo com o modelo multidimensional definido. Esta fase incluiu a seleção de atributos relevantes, a limpeza e normalização dos dados, a conversão de unidades de medida e o carregamento das dimensões e da tabela de factos.

A seleção dos atributos a integrar na *Data Warehouse* foi realizada nesta fase, com base nos objetivos analíticos definidos previamente. Foram considerados essenciais os seguintes atributos:

- **Ano e Mês**, para a dimensão Tempo;
- **Distrito, Concelho e Freguesia**, para a dimensão Localização;

- **Nível de Tensão**, para a dimensão Voltagem;
- **Energia Ativa (kWh)**, como métrica base para o cálculo do consumo elétrico.

Quanto à transformação, foram aplicadas operações de limpeza, normalização e enriquecimento dos dados antes do carregamento na *Data Warehouse*. Foram removidos registo com valores nulos nos atributos essenciais, normalizados os campos textuais através da conversão para minúsculas e eliminados duplicados no carregamento das dimensões.

Na dimensão Tempo, foram derivados os atributos Trimestre, Semestre e Nome do mês, de forma a enriquecer a análise temporal. Na tabela de factos, foi efetuada a conversão da métrica de consumo de kWh para MWh, tendo também sido realizada a conversão e validação do tipo de dados numéricos.

O carregamento da *Data Warehouse* foi realizado através de *stored procedures* específicas para cada dimensão e para a tabela de factos, assegurando a integridade referencial entre as tabelas. As dimensões foram carregadas em primeiro lugar, seguidas da tabela de factos.

7. Exploração Analítica em SQL

Este capítulo apresenta a exploração analítica dos dados armazenados na *Data Warehouse*, recorrendo a diferentes mecanismos de análise disponibilizados pelo SQL Server. O objetivo principal é responder às questões de negócio previamente definidas, demonstrando a capacidade do modelo multidimensional para suportar análises agregadas e multidimensionais.

Para esse efeito, foram utilizados mecanismos estudados em contexto laboratorial, nomeadamente a criação de índices, o uso de *Common Table Expressions* (CTE) e diferentes variantes da cláusula GROUP BY, como ROLLUP, CUBE e GROUPING SETS. Estes mecanismos permitem melhorar o desempenho das consultas e facilitar a análise dos dados sob diferentes perspetivas

7.1. Índices

Os índices são estruturas auxiliares criadas sobre as tabelas de uma base de dados, com o objetivo de melhorar o desempenho das consultas, reduzindo o número de acessos necessários aos dados. No contexto de uma *Data Warehouse*, os índices assumem um papel particularmente importante, uma vez que as consultas analíticas envolvem frequentemente grandes volumes de dados, operações de junção e agregações.

Para além dos índices criados automaticamente aquando da definição das chaves primárias, foram criados índices não clusterizados de forma manual, incidindo sobre as colunas mais utilizadas nas consultas analíticas. Em particular, foram definidos índices nas chaves estrangeiras da tabela de factos, utilizadas nas junções com as dimensões Tempo, Localização e Voltagem, bem como em atributos relevantes das dimensões utilizados em operações de filtragem e agregação.

Embora a aplicação de índices não tenha reduzido significativamente as leituras lógicas em consultas que agregam grande parte da tabela de factos, mantêm-se úteis para junções e filtros seletivos. A eficácia dos índices depende, portanto, do padrão de acesso e do tipo de consulta executada.

Como exemplo, considera-se a seguinte consulta que filtra valores específicos nas dimensões:

```
SELECT SUM(f.consumo)
  FROM dw.fact_consumo_eletrico f
  JOIN dw.dim_tempo t ON f.id_tempo = t.id_tempo
  JOIN dw.dim_localizacao l ON f.id_localizacao = l.id_localizacao
```

```
WHERE l.freguesia = 'Vila Verde' AND t.mes = 12;
```

Nesta consulta, a utilização de índices mostrou um ganho significativo em termos de *logical reads*:

- **Sem índices:** 15.687.840 *logical reads*
- **Com índices:** 10.579.200 *logical reads*
- **Redução/Ganho:** 5.108.640 *logical reads*

Este exemplo evidencia claramente que, quando a *query* envolve filtros seletivos sobre dimensões, os índices permitem localizar rapidamente os registo relevantes na tabela de factos, reduzindo significativamente o número de acessos necessários aos dados e melhorando o desempenho da consulta.

7.2. Common Table Expressions (CTE)

As *Common Table Expressions* (CTE) permitem definir consultas temporárias e reutilizáveis dentro de uma instrução SQL, melhorando a legibilidade e a organização do código. No contexto da exploração analítica da *Data Warehouse*, as CTE foram utilizadas para estruturar consultas agregadas de forma clara e intuitiva.

No contexto deste trabalho, recorreu-se à utilização de uma CTE para demonstrar a aplicação deste mecanismo na estruturação de consultas analíticas, através da análise do consumo elétrico total e médio por distrito e por ano.

A consulta realiza junções entre a tabela de factos e as dimensões Tempo e Localização, aplicando funções de agregação sobre a métrica de consumo. Os resultados obtidos permitem identificar padrões de consumo por distrito e comparar diferentes períodos temporais.

A consulta SQL utilizada encontra-se apresentada em **Anexo**.

7.3. ROLLUP

A cláusula ROLLUP é uma extensão do GROUP BY que permite calcular, numa única consulta, totais e subtotais hierárquicos sobre os dados. No contexto da *Data Warehouse*, este mecanismo é particularmente útil para análises agregadas que envolvem diferentes níveis de detalhe, como totais por categoria e subtotais intermédios.

No âmbito deste trabalho, foi utilizada uma consulta com a cláusula ROLLUP como exemplo representativo da sua aplicação na exploração analítica da *Data Warehouse*. A consulta tem como objetivo analisar o consumo elétrico total por

distrito e por mês, incluindo automaticamente os subtotais por distrito e o total global.

A consulta realiza junções entre a tabela de factos e as dimensões Tempo e Localização, aplicando uma função de agregação sobre a métrica de consumo. A utilização do ROLLUP permite obter uma visão hierárquica do consumo energético, possibilitando a análise comparativa entre diferentes regiões geográficas e períodos temporais, bem como a identificação de padrões mensais e regionais de consumo.

A consulta SQL utilizada encontra-se apresentada em [Anexo](#).

7.4. CUBE

A cláusula CUBE é uma extensão do GROUP BY que permite calcular automaticamente todas as combinações possíveis de agregação entre os atributos definidos. Este mecanismo é particularmente útil em ambientes de *Data Warehousing*, pois possibilita uma análise multidimensional completa sem a necessidade de executar múltiplas consultas.

No contexto deste trabalho, foi utilizada uma consulta com a cláusula CUBE como exemplo representativo da sua aplicação na exploração analítica da *Data Warehouse*. A consulta tem como objetivo analisar o consumo elétrico por distrito e por nível de tensão, permitindo obter simultaneamente os totais por distrito, os totais por nível de tensão, as combinações entre ambos e o total global.

A consulta realiza junções entre a tabela de factos e as dimensões Localização e Voltagem, aplicando uma função de agregação sobre a métrica de consumo. A utilização do CUBE permite identificar padrões de consumo associados aos diferentes níveis de tensão em cada região geográfica, bem como comparar o impacto dos diferentes níveis de voltagem no consumo energético.

A consulta SQL utilizada encontra-se apresentada em [Anexo](#).

7.5. GROUPING SETS

A cláusula GROUPING SETS permite definir explicitamente diferentes níveis de agregação numa única consulta, oferecendo maior flexibilidade e controlo sobre os resultados obtidos quando comparada com outras extensões do GROUP BY. Este mecanismo é particularmente útil quando se pretende analisar os dados sob várias perspetivas específicas, sem recorrer à execução de múltiplas consultas independentes.

Neste trabalho, foi utilizada uma consulta com a cláusula GROUPING SETS como exemplo representativo da sua aplicação na exploração analítica da *Data*

Warehouse. A consulta tem como objetivo analisar o consumo elétrico por ano e por nível de voltagem, permitindo obter simultaneamente os totais por combinação ano/voltagem, os totais por ano, os totais por nível de voltagem e o total global.

A consulta realiza junções entre a tabela de factos e as dimensões Tempo e Voltagem, aplicando uma função de agregação sobre a métrica de consumo. A utilização dos GROUPING SETS possibilita uma análise clara da evolução temporal do consumo energético, bem como da distribuição do consumo pelos diferentes níveis de tensão, permitindo comparar diferentes perspetivas analíticas numa única consulta.

A consulta SQL utilizada encontra-se apresentada no **Anexo**.

8. Análise em Excel (PowerPivot)

Complementarmente à análise em SQL, foi utilizada a ferramenta Microsoft Excel com PowerPivot, permitindo a exploração dos dados da *Data Warehouse* de forma interativa. A ligação entre o Excel e a *Data Warehouse* possibilitou a criação de tabelas dinâmicas baseadas no modelo multidimensional, facilitando a análise do consumo segundo diferentes perspetivas.

Através do PowerPivot, foi possível responder às mesmas questões de negócio definidas anteriormente, nomeadamente o consumo por distrito e ano, por distrito e mês, por nível de tensão e por ano e nível de voltagem. Esta abordagem permite uma análise mais visual e dinâmica, possibilitando a aplicação de filtros, segmentações e ordenações de forma intuitiva, sem necessidade de escrever consultas SQL.

8.1. Consumo por Distrito e Ano

	A	B	C	D	E	F
1						
2						
Qual é o consumo por distrito e ano?						
4	Rótulos de Linha	Média de consumo	Soma de consumo			
5	aveiro	1387,843979	20986976,64			
6	beja					
7	2020	601,4112889	162381,048			
8	2021	665,0316092	1087326,681			
9	2022	690,5363204	1135932,247			
10	2023	729,1554036	1199460,639			
11	2024	701,6024863	1155539,295			
12	2025	710,4925602	785094,279			
13	braga	526,5298608	16807886,22			
14	bragança	127,5357777	1901048,302			
15	castelo branco	399,14045	3991404,5			
16	coimbra	923,8724336	12633955,53			
17	évora	481,548926	3683367,735			
18	faro	1603,862892	11563851,45			
19	guarda	133,8987658	2343763,996			
20	leiria	1023,854644	12351782,42			
21	lisboa	2822,110425	42278036,28			
22	portalegre	325,4123345	2315634,172			
23	porto	1486,877866	36198041,65			
24	santarém	715,5368567	10429665,22			
25	setúbal	3852,310064	23930550,12			
26	viana do castelo	348,5328605	5418988,915			
27	vila real	178,0274795	2792539,044			
28	viseu	292,5616139	6744130,323			
29	Total Geral	855,0463623	221897356,7			
30						
31						

Figura 4 - PowerPivot Consumo por Distrito e Ano

A **Figura 4** apresenta a análise do consumo de energia por distrito ao longo dos diferentes anos, obtida através de uma tabela dinâmica construída em Excel com recurso ao PowerPivot.

8.2. Consumo total por Distrito e Mês

A	B	C	D	E	F	G	H	I	J	K
1										
2	Qual é consumo total por distrito e mês, incluindo subtotais?									
3										
4	Rótulos de Linha	Soma de consumo								
5	aveiro									
6	January	2004989,512								
7	February	1812699,013								
8	March	1918415,645								
9	April	1737022,149								
10	May	1801094,737								
11	June	1747342,731								
12	July	1805341,836								
13	August	1545606,696								
14	September	1410418,618								
15	October	1494975,427								
16	November	1885940,888								
17	December	1823129,392								
18	beja	5525734,189								
19	braga	16807886,22								
20	bragança	1901048,302								
21	castelo branco	3991404,5								
22	coimbra	12633955,53								
23	évora	3683367,735								
24	faro	11563851,45								
25	guarda	2343763,996								
26	leiria	12351782,42								
27	lisboa	42278036,28								
28	portalegre	2315634,172								
29	porto	36198041,65								
30	santarém	10429665,22								
31	setúbal	23930550,12								
32	viana do castelo	5418988,915								
33	vila real	2792539,044								
34	viseu	6744130,323								
35	Total Geral	221897356,7								

Figura 5 - PowerPivot Consumo total por Distrito e Mês

A **Figura 5** apresenta a análise do consumo de energia por distrito ao longo dos diferentes meses, obtida através de uma tabela dinâmica construída em Excel com recurso ao PowerPivot.

8.3. Consumo por Distrito e Nível de Tensão

A	B	C	D	E	F
Qual é o consumo por distrito e nível de tensão?					
	Soma de consumo	Rótulos de Coluna			
	Rótulos de Linha	baixa tensão	muito alta, alta e média tensões	Total Geral	
aveiro	6813026,912		14173949,73	20986976,64	
beja	1730284,333		3795449,856	5525734,189	
braga	8121028,74		8686857,475	16807886,22	
bragança	1480842,228		420206,074	1901048,302	
castelo branco	2144778,698		1846625,802	3991404,5	
coimbra	4271877,763		8362077,766	12633955,53	
évora	1949464,153		1733903,582	3683367,735	
faro	8436855,644		3126995,804	11563851,45	
guarda	1549917,642		793846,354	2343763,996	
leiria	5311606,357		7040176,065	12351782,42	
lisboa	23825975,13		18452061,15	42278036,28	
portalegre	1327795,013		987839,159	2315634,172	
porto	18928395,5		17269646,14	36198041,65	
santarém	4745267,033		5684398,19	10429665,22	
setúbal	8291752,163		15638797,95	23930550,12	
viana do castelo	2538468,425		2880520,49	5418988,915	
vila real	1933311,181		859227,863	2792539,044	
viseu	3503497,312		3240633,011	6744130,323	
Total Geral	106904144,2		114993212,5	221897356,7	

Figura 6 - PowerPivot Consumo total por Distrito e Nível de Tensão

A **Figura 6** ilustra o consumo total de energia por distrito, com discriminação dos diferentes níveis de tensão.

8.4. Consumo por Ano e Nível de Tensão

	A	B	C	D	E
1					
2		Consumo por Ano e Nível de Tensão?			
3					
4	Soma de consumo	Rótulos de Coluna			
5	Rótulos de Linha	baixa tensão	muito alta, alta e média tensões	Total Geral	
6	2020	3885484,073	3851237,472	7736721,545	
7	2021	21102472,76	23477990,22	44580462,97	
8	2022	21627682,68	23931398,64	45559081,32	
9	2023	22108439,42	23674295,86	45782735,28	
10	2024	22468446,45	24085312,82	46553759,28	
11	2025	15711618,85	15972977,46	31684596,31	
12	Total Geral	106904144,2	114993212,5	221897356,7	
13					
14					
15					
16					
17					
18					

Figura 7 - PowerPivot Consumo total por Ano e Nível de Tensão

A **Figura 7** apresenta a evolução do consumo de energia por ano, discriminada por nível de voltagem.

9. Visualização em PowerBI

Por fim, foi utilizada a ferramenta Power BI como plataforma de visualização dos dados, permitindo a criação de relatórios e *dashboards* interativos. Foram utilizados os visuais disponibilizados por defeito, como gráficos de barras, colunas e tabelas, bem como pelo menos um visual adicional, de forma a enriquecer a apresentação dos resultados.

O Power BI permitiu representar graficamente as análises efetuadas, facilitando a interpretação dos padrões de consumo energético e tornando os resultados mais acessíveis para utilizadores não técnicos. Esta ferramenta complementa as análises realizadas em SQL e Excel, evidenciando o papel da *Data Warehouse* como base central para diferentes soluções.

9.1. Consumo por Distrito e Ano

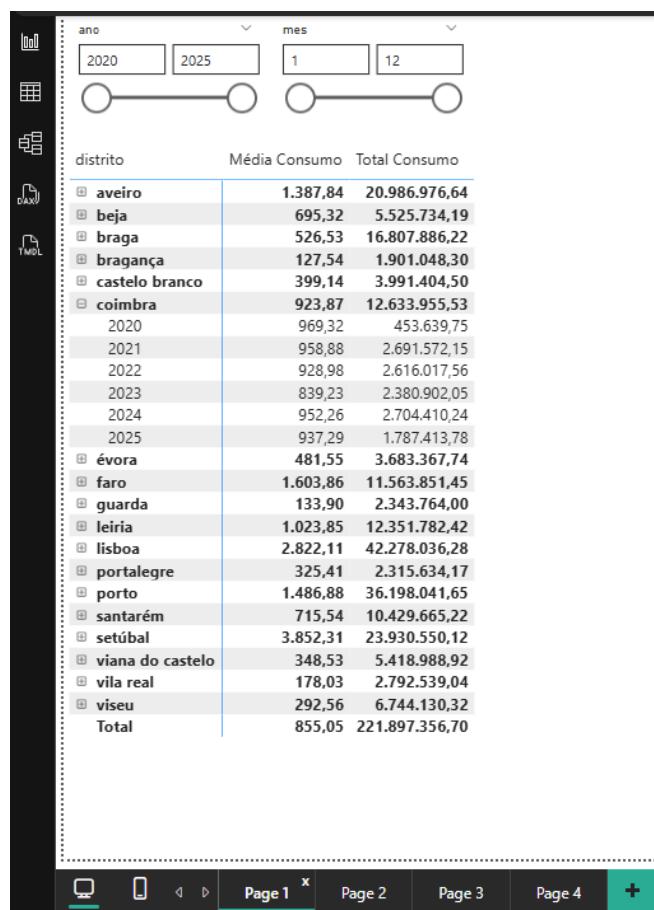


Figura 8 - PowerBI Consumo por Distrito e Ano

A **Figura 8** apresenta a análise do consumo de energia por distrito ao longo dos diferentes anos, obtida da criação de uma matriz em PowerBI.

9.2. Consumo total por Distrito e Mês

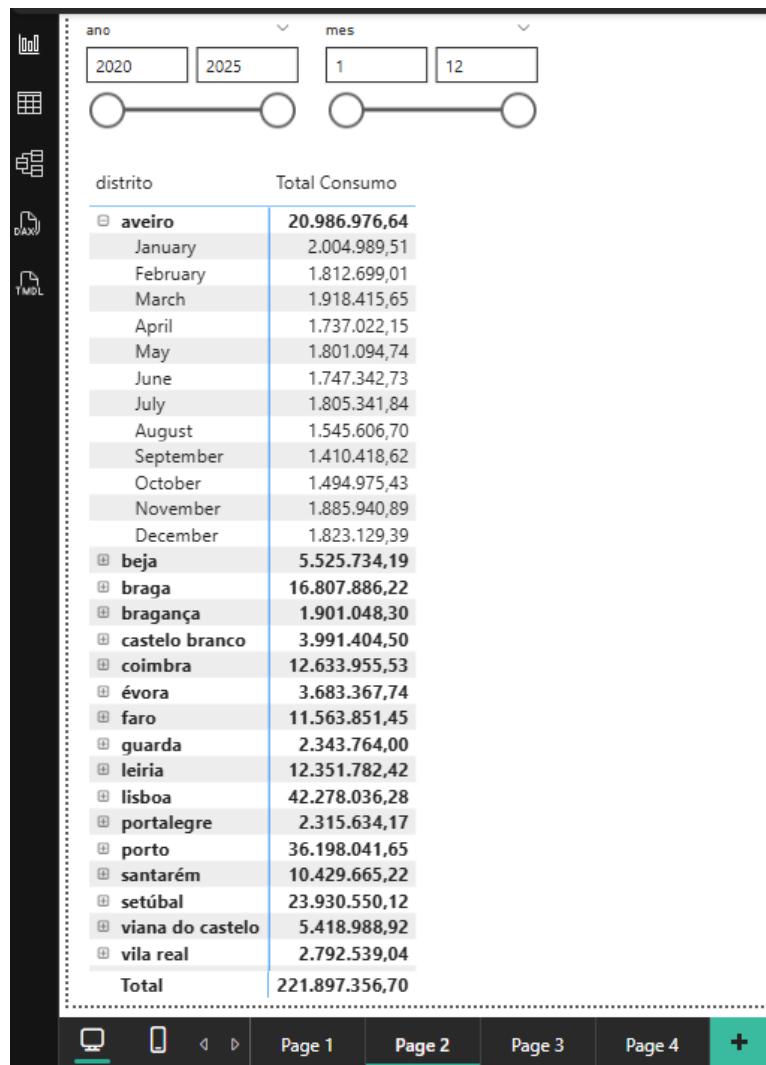


Figura 9 - PowerBI Consumo total por Distrito e Mês

A **Figura 9** apresenta a análise do consumo de energia por distrito ao longo dos diferentes meses.

9.3. Consumo por Distrito e Nível de Tensão

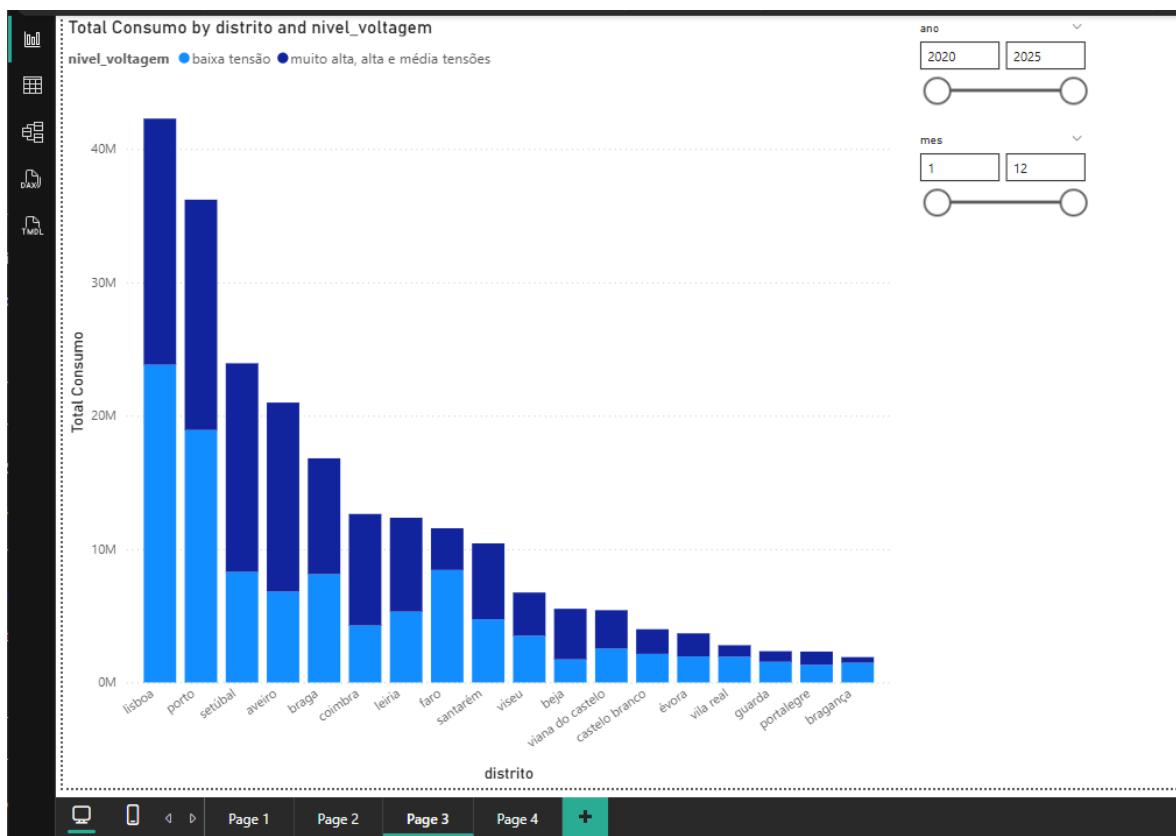


Figura 10 - PowerBI Consumo total por Distrito e Nível de Tensão

A **Figura 10** ilustra o consumo total de energia por distrito, com discriminação dos diferentes níveis de tensão.

9.4. Consumo por Ano e Nível de Tensão

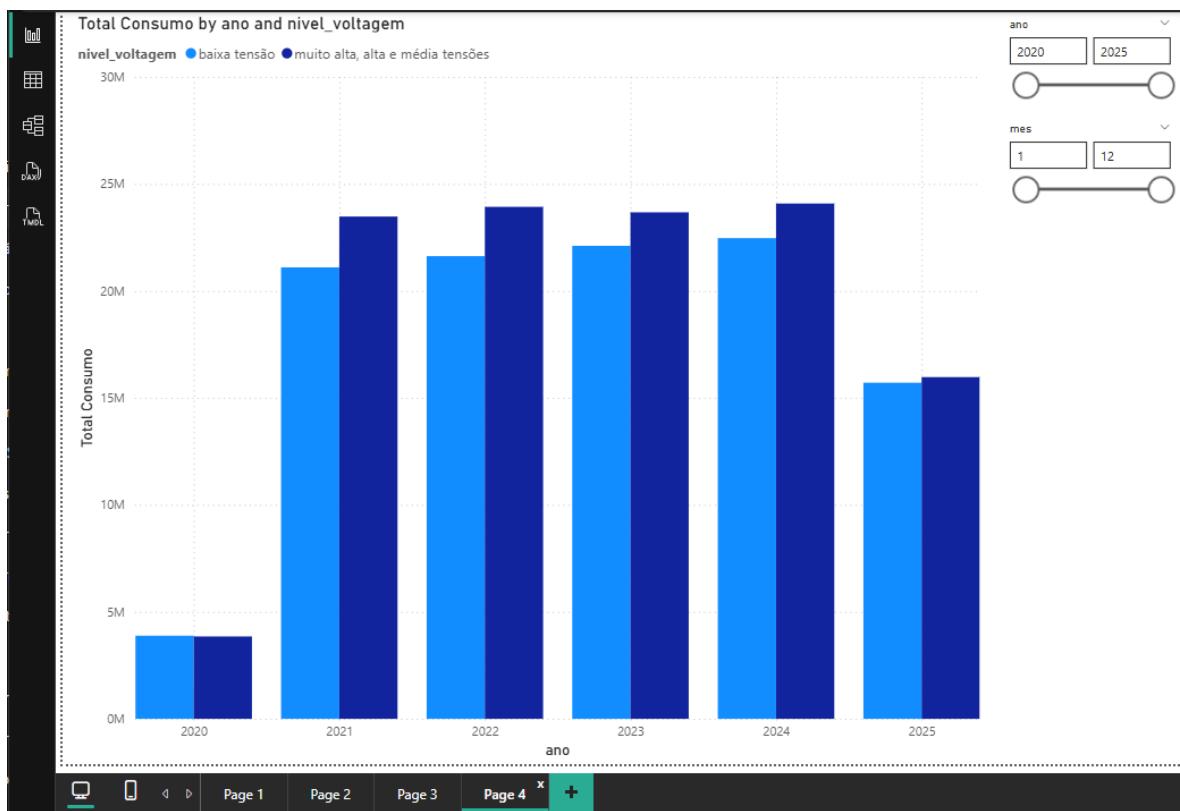


Figura 11 - PowerBI Consumo total por Ano e Nível de Tensão

A **Figura 11** apresenta a evolução do consumo de energia por ano, discriminada por nível de voltagem.

10. Relatórios de Apoio à Decisão?

As margens do documento deverão respeitar as seguintes medidas: superior 3 cm, inferior 2 cm, interior 3 cm, exterior 2,5 cm.

11. Conclusão

Em conclusão, o desenvolvimento deste trabalho permitiu consolidar, de forma prática, os conhecimentos adquiridos ao longo da unidade curricular, evidenciando a importância dos sistemas de *Data Warehousing* no apoio à análise e à tomada de decisão. A implementação de uma *Data Warehouse* possibilitou uma melhor compreensão dos princípios da modelação multidimensional e do impacto que uma estrutura adequada tem na eficiência e flexibilidade da análise dos dados.

O contacto com ferramentas como o Excel com PowerPivot e o Power BI contribuiu para uma melhor percepção do valor da análise visual e interativa dos dados.

Adicionalmente, este trabalho possibilitou não só a aplicação prática dos conceitos teóricos abordados, mas também o desenvolvimento de uma visão mais crítica sobre o papel dos dados no contexto organizacional, reforçando a importância de soluções analíticas bem estruturadas para a tomada de decisões informadas.

12. Referências

- [1] “Explore - E-REDES Open Data Portal - Huwise,” E-Redes, [Online]. Available: <https://e-redes.opendatasoft.com/explore/?sort=modified>. [Acedido em 26 Dezembro 2026].
- [2] “Consumos mensais por concelho,” E-Redes, 20 Novembro 2025. [Online]. Available: <https://e-redes.opendatasoft.com/explore/dataset/3-consumos-faturados-por-municipio-ultimos-10-anos/information/>. [Acedido em 26 Dezembro 2025].

13. Anexo

O Capítulo de Anexo apresenta todo o código SQL desenvolvido no decorrer deste trabalho estando este dividido em 4 ficheiros, o ficheiro dos *Drops* para limpar toda a base de dados e possibilitar a execução limpa dos outros ficheiros, o ficheiro relativo à criação da estrutura da *Data Warehouse*, o ficheiro de criação e execução das *procedures* responsáveis pela transformação e carregamento dos dados e por ultimo o ficheiro com o código relativo às *queries* feitas que respondem às perguntas definidas sobre o objetivo desta *Data Warehouse*.

13.1. Drops

```
USE dw8_2526_DW_Energia;
```

----- Drops -----

```
-- remover procedures
DROP PROCEDURE IF EXISTS dw.sp_popular_dim_localizacao;
DROP PROCEDURE IF EXISTS dw.sp_popular_dim_tempo;
DROP PROCEDURE IF EXISTS dw.sp_popular_dim_voltagem;
DROP PROCEDURE IF EXISTS dw.sp_popular_fact_consumo;
-- remover tabelas
DROP TABLE IF EXISTS dw.fact_consumo_eletrico;
DROP TABLE IF EXISTS dw.dim_tempo;
DROP TABLE IF EXISTS dw.dim_voltagem;
DROP TABLE IF EXISTS dw.dim_localizacao;
-- remover partições
-- Dropar Partition Scheme se existir
IF EXISTS (SELECT * FROM sys.partition_schemes WHERE name = 'ps_fact_por_id_tempo')
    DROP PARTITION SCHEME ps_fact_por_id_tempo;
-- Dropar Partition Function se existir
```

```

    IF EXISTS (SELECT * FROM sys.partition_functions WHERE name =
'pf_fact_por_id_tempo')

        DROP PARTITION FUNCTION pf_fact_por_id_tempo;

    -- remover ficheiros (só se existirem)

    IF EXISTS (SELECT 1 FROM sys.database_files WHERE name =
'dw_dim_20221849')

        ALTER DATABASE dw8_2526_DW_Energia REMOVE FILE
dw_dim_20221849;

    IF EXISTS (SELECT 1 FROM sys.database_files WHERE name =
'dw_pk_20221849')

        ALTER DATABASE dw8_2526_DW_Energia REMOVE FILE
dw_pk_20221849;

    IF EXISTS (SELECT 1 FROM sys.database_files WHERE name =
'dw_fact_1_20221849')

        ALTER DATABASE dw8_2526_DW_Energia REMOVE FILE
dw_fact_1_20221849;

    IF EXISTS (SELECT 1 FROM sys.database_files WHERE name =
'dw_fact_2_20221849')

        ALTER DATABASE dw8_2526_DW_Energia REMOVE FILE
dw_fact_2_20221849;

    -- remover filegroups (só se existirem)

    IF EXISTS (SELECT 1 FROM sys.filegroups WHERE name = 'FG_DIM')

        ALTER DATABASE dw8_2526_DW_Energia REMOVE FILEGROUP
FG_DIM;

    IF EXISTS (SELECT 1 FROM sys.filegroups WHERE name = 'FG_PK')

        ALTER DATABASE dw8_2526_DW_Energia REMOVE FILEGROUP
FG_PK;

    IF EXISTS (SELECT 1 FROM sys.filegroups WHERE name = 'FG_FACT_1')

        ALTER DATABASE dw8_2526_DW_Energia REMOVE FILEGROUP
FG_FACT_1;

    IF EXISTS (SELECT 1 FROM sys.filegroups WHERE name = 'FG_FACT_2')

        ALTER DATABASE dw8_2526_DW_Energia REMOVE FILEGROUP
FG_FACT_2;

    -- remover schema dw

    IF EXISTS (SELECT 1 FROM sys.schemas WHERE name = 'dw')

```

```
EXEC('DROP SCHEMA dw');

-- Remover índices da fact table se existirem
IF EXISTS (SELECT 1
            FROM sys.indexes
              WHERE name = 'idx_fact_time' AND object_id =
OBJECT_ID('dw.fact_consumo_eletrico'))
    DROP INDEX idx_fact_time ON dw.fact_consumo_eletrico;
IF EXISTS (SELECT 1
            FROM sys.indexes
              WHERE name = 'idx_fact_location' AND object_id =
OBJECT_ID('dw.fact_consumo_eletrico'))
    DROP INDEX idx_fact_location ON dw.fact_consumo_eletrico;
IF EXISTS (SELECT 1
            FROM sys.indexes
              WHERE name = 'idx_fact_voltage' AND object_id =
OBJECT_ID('dw.fact_consumo_eletrico'))
    DROP INDEX idx_fact_voltage ON dw.fact_consumo_eletrico;
-- Remover índices das dimensões se existirem
IF EXISTS (SELECT 1
            FROM sys.indexes
              WHERE name = 'idx_dim_location' AND object_id =
OBJECT_ID('dw.dim_localizacao'))
    DROP INDEX idx_dim_location ON dw.dim_localizacao;
IF EXISTS (SELECT 1
            FROM sys.indexes
              WHERE name = 'idx_dim_time' AND object_id =
OBJECT_ID('dw.dim_tempo'))
    DROP INDEX idx_dim_time ON dw.dim_tempo;
```

13.2. Criar Data Warehouse

```
USE dw8_2526_DW_Energia;
```

```
-----  
----- Create Schema -----  
-----
```

```
IF NOT EXISTS (SELECT 1 FROM sys.schemas WHERE name = 'dw')  
    EXEC('CREATE SCHEMA dw');
```

```
-----  
----- Particoes -----  
-----
```

```
ALTER DATABASE dw8_2526_DW_Energia ADD FILEGROUP FG_DIM;  
ALTER DATABASE dw8_2526_DW_Energia ADD FILEGROUP FG_PK;  
ALTER DATABASE dw8_2526_DW_Energia ADD FILEGROUP FG_FACT_1;  
ALTER DATABASE dw8_2526_DW_Energia ADD FILEGROUP FG_FACT_2;
```

```
ALTER DATABASE dw8_2526_DW_Energia ADD FILE (  
    NAME = dw_dim_20221849,  
    FILENAME = 'f:\fg\dw_dim_20221849.ndf',  
    SIZE = 10MB,  
    FILEGROWTH = 10MB  
) TO FILEGROUP FG_DIM;
```

```
ALTER DATABASE dw8_2526_DW_Energia ADD FILE (  
    NAME = dw_pk_20221849,  
    FILENAME = 'g:\fg\dw_pk_20221849.ndf',  
    SIZE = 10MB,  
    FILEGROWTH = 10MB  
) TO FILEGROUP FG_PK;
```

```
ALTER DATABASE dw8_2526_DW_Energia ADD FILE (
    NAME = dw_fact_1_20221849,
    FILENAME = 'h:\fgh\dw_fact_1_20221849.ndf',
    SIZE = 20MB,
    FILEGROWTH = 20MB
) TO FILEGROUP FG_FACT_1;
```

```
ALTER DATABASE dw8_2526_DW_Energia ADD FILE (
    NAME = dw_fact_2_20221849,
    FILENAME = 'i:\fgi\dw_fact_2_20221849.ndf',
    SIZE = 20MB,
    FILEGROWTH = 20MB
) TO FILEGROUP FG_FACT_2;
```

----- Create Tables -----

```
CREATE TABLE dw.dim_tempo (
    id_tempo INT NOT NULL,
    ano INT NOT NULL,
    mes INT NOT NULL,
    trimestre INT NOT NULL,
    semestre INT NOT NULL,
    nome_mes VARCHAR(20)
) ON FG_DIM;

ALTER TABLE dw.dim_tempo
ADD CONSTRAINT pk_dim_tempo
PRIMARY KEY (id_tempo)
ON FG_PK;
```

```
CREATE TABLE dw.dim_localizacao (
    id_localizacao INT NOT NULL IDENTITY(1,1),
    distrito VARCHAR(200),
    concelho VARCHAR(200),
    freguesia VARCHAR(200)
) ON FG_DIM;
```

```
ALTER TABLE dw.dim_localizacao
ADD CONSTRAINT pk_dim_localizacao
PRIMARY KEY (id_localizacao)
ON FG_PK;
```

```
CREATE TABLE dw.dim_voltagem (
    id_voltagem INT NOT NULL IDENTITY(1,1),
    nivel_voltagem VARCHAR(50)
) ON FG_DIM;

ALTER TABLE dw.dim_voltagem
ADD CONSTRAINT pk_dim_voltagem
PRIMARY KEY (id_voltagem)
ON FG_PK;
```

```
-- Partition function usando id_tempo
CREATE PARTITION FUNCTION pf_fact_por_id_tempo(INT)
AS RANGE LEFT
FOR VALUES (0); -- tudo <=0 vai para FG_FACT_1, >0 vai para FG_FACT_2
```

```
-- Partition scheme que associa a partition function aos filegroups
CREATE PARTITION SCHEME ps_fact_por_id_tempo
AS PARTITION pf_fact_por_id_tempo
TO (FG_FACT_1, FG_FACT_2);
```

```

CREATE TABLE dw.fact_consumo_eletrico (
    id_consumo_eletrico BIGINT NOT NULL IDENTITY(1,1),
    consumo DECIMAL(12,3),
    id_tempo INT NOT NULL,
    id_localizacao INT NOT NULL,
    id_voltagem INT NOT NULL,
    CONSTRAINT pk_fact_consumo PRIMARY KEY CLUSTERED
    (id_consumo_eletrico, id_tempo)
) ON ps_fact_por_id_tempo(id_tempo);

-- FKs continuam iguais
ALTER TABLE dw.fact_consumo_eletrico
ADD CONSTRAINT fk_fact_tempo
FOREIGN KEY (id_tempo)
REFERENCES dw.dim_tempo(id_tempo);

ALTER TABLE dw.fact_consumo_eletrico
ADD CONSTRAINT fk_fact_localizacao
FOREIGN KEY (id_localizacao)
REFERENCES dw.dim_localizacao(id_localizacao);

ALTER TABLE dw.fact_consumo_eletrico
ADD CONSTRAINT fk_fact_voltagem
FOREIGN KEY (id_voltagem)
REFERENCES dw.dim_voltagem(id_voltagem);

-- Índices na fact table e dimensões
CREATE NONCLUSTERED INDEX idx_fact_time
ON dw.fact_consumo_eletrico (id_tempo);
CREATE NONCLUSTERED INDEX idx_fact_location

```

```
ON dw.fact_consumo_eletrico (id_localizacao);
CREATE NONCLUSTERED INDEX idx_fact_voltage
ON dw.fact_consumo_eletrico (id_voltagem);
CREATE NONCLUSTERED INDEX idx_dim_location
ON dw.dim_localizacao (distrito, concelho, freguesia);
CREATE NONCLUSTERED INDEX idx_dim_time
ON dw.dim_tempo (ano, mes);
```

13.3. Transformação e Carregamento

```
USE dw8_2526_DW_Energia;
```

```
----- Procedure dim_localizacao -----
```

```
CREATE PROCEDURE dw.sp_popular_dim_localizacao
AS
BEGIN
    INSERT INTO dw.dim_localizacao (distrito, concelho, freguesia)
    SELECT DISTINCT
        LOWER(Distrito),
        LOWER(Concelho),
        LOWER(Freguesia)
    FROM staging.consumo_energia_raw
    WHERE Distrito IS NOT NULL
        AND Concelho IS NOT NULL
        AND Freguesia IS NOT NULL;
END;
```

```
----- Procedure dim_tempo -----
```

```
CREATE PROCEDURE dw.sp_popular_dim_tempo
AS
BEGIN
    INSERT INTO dw.dim_tempo (
        id_tempo,
```

```

ano,
mes,
trimestre,
semestre,
nome_mes

)
SELECT
CASE
    WHEN Ano <= 2022 THEN -ROW_NUMBER() OVER (ORDER BY
Ano, Mês) -- IDs negativos para <= 2022
    WHEN Ano >= 2023 THEN ROW_NUMBER() OVER (ORDER BY Ano,
Mês) -- IDs positivos para >= 2023
    END AS id_tempo,
Ano,
Mês,
DATEPART(QUARTER, DATEFROMPARTS(Ano, Mês, 1)) AS trimestre,
CASE
    WHEN Mês <= 6 THEN 1
    ELSE 2
    END AS semestre,
DATENAME(MONTH, DATEFROMPARTS(Ano, Mês, 1)) AS nome_mes
FROM (
    SELECT DISTINCT Ano, Mês
    FROM staging.consumo_energia_raw
    WHERE Ano IS NOT NULL AND Mês IS NOT NULL
) AS t;
END;

```

----- Procedure dim voltagem -----

```

CREATE PROCEDURE dw.sp_popular_dim_voltagem
AS
BEGIN
    INSERT INTO dw.dim_voltagem (nivel_voltagem)
    SELECT DISTINCT
        LOWER([Nível_de_Tensão])
    FROM staging.consumo_energia_raw
    WHERE [Nível_de_Tensão] IS NOT NULL;
END;

```

----- Procedure Fact table -----

```

CREATE PROCEDURE dw.sp_popular_fact_consumo
AS
BEGIN
    INSERT INTO dw.fact_consumo_eletrico (
        consumo,
        id_tempo,
        id_localizacao,
        id_voltagem
    )
    SELECT
        -- Conversão de kWh para MWh
        TRY_CAST(REPLACE(s.[Energia_Ativa_kWh], ',', '.') AS DECIMAL(12,3))
        / 1000.0 AS consumo_mwh,
        t.id_tempo,
        l.id_localizacao,
        v.id_voltagem

```

```
FROM staging.consumo_energia_raw s
JOIN dw.dim_tempo t
    ON t.ano = s.Ano AND t.mes = s.Mês
JOIN dw.dim_localizacao l
    ON l.distrito = LOWER(s.Distrito)
    AND l.concelho = LOWER(s.Concelho)
    AND l.freguesia = LOWER(s.Freguesia)
JOIN dw.dim_voltagem v
    ON v.nivel_voltagem = LOWER(s.[Nível_de_Tensão])
WHERE s.[Energia_Ativa_kWh] IS NOT NULL;
END;
```

```
-- Popular Tabelas
EXEC dw.sp_popular_dim_tempo;
EXEC dw.sp_popular_dim_localizacao;
EXEC dw.sp_popular_dim_voltagem;
EXEC dw.sp_popular_fact_consumo;
```

13.4. Queries SQL

```
USE dw8_2526_DW_Energia;
```

```
SET STATISTICS IO ON;
```

```
SET STATISTICS TIME ON;
```

```
-- Indice (Valores específicos dims)
```

```
SELECT SUM(f.consumo)
```

```
FROM dw.fact_consumo_eletrico f
```

```
JOIN dw.dim_tempo t ON f.id_tempo = t.id_tempo
```

```
JOIN dw.dim_localizacao l ON f.id_localizacao = l.id_localizacao
```

```
WHERE l.freguesia = 'Vila Verde' AND t.mes = 12;
```

```
-- LR sem Indices ->580*1288*21 = 15687840
```

```
-- LR com Indices ->580*912*20 = 10579200
```

```
-- Ganho -> 15687840 - 10579200 = 5108640
```

```
-- CTE: Qual é o consumo por distrito e ano?
```

```
WITH ConsumoPorDistritoAno AS (
```

```
    SELECT
```

```
        l.distrito,
```

```
        t.ano,
```

```
        SUM(f.consumo) AS Total_Consumo,
```

```
        AVG(f.consumo) AS Media_Consumo
```

```
    FROM dw.fact_consumo_eletrico f
```

```
    JOIN dw.dim_localizacao l ON f.id_localizacao = l.id_localizacao
```

```
    JOIN dw.dim_tempo t ON f.id_tempo = t.id_tempo
```

```

        GROUP BY l.distrito, t.ano
    )
SELECT *
FROM ConsumoPorDistritoAno
ORDER BY distrito, ano;

```

-- ROLLUP: Qual é consumo total por distrito e mês, incluindo subtotais?

```

SELECT
    l.distrito,
    t.mes,
    SUM(f.consumo) AS Total_Consumo
FROM dw.fact_consumo_eletrico f
JOIN dw.dim_localizacao l ON f.id_localizacao = l.id_localizacao
JOIN dw.dim_tempo t ON f.id_tempo = t.id_tempo
GROUP BY ROLLUP (l.distrito, t.mes)
ORDER BY l.distrito, t.mes;

```

--CUBE: Qual é o consumo por distrito e nível de tensão?

```

SELECT
    l.distrito,
    v.nivel_voltagem AS Voltage_Level,
    SUM(f.consumo) AS Total_Consumo
FROM dw.fact_consumo_eletrico f
JOIN dw.dim_localizacao l ON f.id_localizacao = l.id_localizacao
JOIN dw.dim_voltagem v ON f.id_voltagem = v.id_voltagem
GROUP BY CUBE (l.distrito, v.nivel_voltagem)
ORDER BY l.distrito, v.nivel_voltagem;

```

-- GROUPING SETS: Consumo por Ano e Nível de tensão?

SELECT

```
t.ano,  
v.nivel_voltagem,  
SUM(f.consumo) AS Total_Consumo  
FROM dw.fact_consumo_eletrico f  
JOIN dw.dim_tempo t  
    ON f.id_tempo = t.id_tempo  
JOIN dw.dim_voltagem v  
    ON f.id_voltagem = v.id_voltagem  
GROUP BY GROUPING SETS (  
    (t.ano, v.nivel_voltagem),  
    (t.ano),  
    (v.nivel_voltagem),  
    ()  
)  
ORDER BY t.ano, v.nivel_voltagem;
```