Data warehousing é o processo de recolher e gerir informação de várias fontes, com o objectivo de conseguir percepções de negócio significativas. Este processo tambem permite o data mining, que é a procura de padrões em dados guardados numa base de dados, de forma a tentar obter mais vendas e lucro.

Um data warehouse é um tipo de sistema de gestão de informação que conecta e analisa informação de várias fontes, e está desenhado para permitir e apoiar atividades de business inteligence. Os Data Warehouses contêm muitos dados, e o seu objectivo é executar operações para fins analíticos e de relatórios de negócios, algo que ajuda a manter registos anteriores e analizar dados para optimizar as operações de negócio.

A diferença entre uma simples base de dados SQL e um data warehouse é que enquanto um apenas consiste numa tecnica de armazenamento de dados, o ultimo está expecialmente projectado para analise dos dados que contem.

Como é que funciona?

Um Data Warehouse funciona como um repositório central em que a informação provém de uma ou mais fontes de dados. Os dados fluem pelo data warehouse via o seu sistema transacional e outras base de dados relacionais.

Os dados que entram num data warehouse podem ser:

* Estruturados;
* Semi-Estruturados;
* Não Estruturados.

Os dados são processados, transformados e consumidos de forma a permitir os utilizadores acederem aos mesmos via ferramentas de BI, clientes SQL, e até excel spreadsheets. Um Data Warehouse junta informação de várias fontes e de todos os tipos, numa só base de dados.

Graças a esta junção, uma empresa consegue analizar os seus clientes de uma forma mais abrangente, ao garantir que toda a informação disponivel é considerada.

Quem é que precisa de um data warehouse?

Em suma, aqueles que necessitam de uma tecnologia de armazenamento e gestão de uma vasta quantidade informação, que torne relativamente simples e rápido o acesso aos dados.

Alguns dos sectores que usufruem de data warehouses são:

* O sector Aéreo: Usa data warehouses para efectuar analises de lucro em rótas, promoções de programas de passageiros frequentes,etc…
* O sector bancário: Usa data warehouses para gerir e optimizar a entrega de recursos como informações dos clientes. Alguns bancos utilizam data warehouses para fazer pesquisas de mercado, analises de performance de productos e operações.
* O sector da saúde: Utiliza data warehouses para planear e prever resultados, gerar diagnosticos de tratamento de pacientes, partilhar dados com companias de seguros, serviços de ajuda médica…

[What is Data Warehouse? Types, Definition & Example (guru99.com)](https://www.guru99.com/data-warehousing.html)

Data Mart

Ambos os data marts e os data warehouses são repositórios estruturados onde dados são armazenados e geridos quanto for necessário. No entanto eles distinguem-se no tipo de informação que contêm:

Os Data warehouses tipicamente são usados como o armazem central de dados de um negócio inteiro. No entanto os data marts só servem uma dada divisão ou função de negócio. Pesquisar informação via um data warehouse consegue ser uma operação muito custuosa para o negócio, e portanto os data marts servem para isolar um pedaço de dados de um inteiro para melhor permitir o acesso dos clientes aos mesmos.

Um data mart pode ser criado a partir de um data warehouse existente (top-down), ou um negócio pode escolher separar unidades de negócio e criar data marts baseando-se nos seus requisitos. Vários data marts podem ser conectados para criar um único data warehouse(bottom-up).

[What is a Data Mart? (vs a Data Warehouse) - Talend | Talend](https://www.talend.com/resources/what-is-data-mart/)

ETL

ETL (sigla para Extrai, Transforma e Carrega), é um processo no Data Warehousing em que uma ferramenta extrai informação de várias fontes de dados, transforma-a e carrega-a para o sistema de data warehouse.

As etapas do ETL podem ser compreendidas da seguinte forma:

1. Extração: Neste passo são obtidas informações vários tipos e de fontes diferentes, e subsequentemente colocadas na area de transformação.
2. Transformação: Neste passo as informações obtidas são convertidas para um formato expecifico. Durante esta etapa podem ocorrer varias operações sobre as informações como a filtragem, limpeza, ordenação…
3. Carregamento: Durante este passo, o resultado da transformação da informação é finalmente carregado para o data warehouse.

É importante notar que o processo ETL pode usar o conceito de pipeline, ou seja, enquanto alguma informação encontra-se no processo de transformação, nova informação pode ser extraida em paralelo.

[ETL Process in Data Warehouse - GeeksforGeeks](https://www.geeksforgeeks.org/etl-process-in-data-warehouse/)