

UNIVERSIDADE DE BRASÍLIA

Faculdade do Gama

Sistemas de Banco de Dados 2

Tecnologias de Banco de Dados (TI-BD)

Data Warehouse e Data Marts

Wesley Lira Carvalho - 200044559

Brasília, DF

2025

1. Definição

A tecnologia em estudo nasceu em meados da década de 1990, onde havia um aumento nos sistemas de gestão empresarial, que por consequência, o aumento também na quantidade de dados. De acordo com RASLAN e CALAZANS (2014), foi necessário que os bancos de dados recorressem a evoluir para que pudessem responder ao crescimento tecnológico e a atmosfera de gestão informatizada, onde nesse novo contexto era necessário trabalhar com dados distintos para uni-los externamente. (RASLAN; CALAZANS, 2014)

O Data Warehouse (DW) é um sistema que agrega diversos dados de diferentes fontes para um único repositório central e consistente, que no qual permite que uma organização possa executar análises fortes em quantidades enormes de dados, na qual um banco de dados comum não pode. Possibilita que a organização dos dados, sendo de clientes, fornecedores e operações, possa obter as informações estratégicas por meio das análises de dados sob novas perspectivas da empresa. Dessa forma, o DW organiza as informações que veem da organização, essas que nos quais a maioria são fontes heterogêneas, filtrando, limpando e preparando os dados para a análise e suporte à decisão. (RASLAN; CALAZANS, 2014)

Em relação aos dados inseridos em uma Data Warehouse, eles passam por um processo de filtragem, limpeza e sumarização, que no qual é chamada de *Data Warehousing*. O processo de Data Warehousing resume-se na retirada dos dados de origem dos sistemas transacionais e também da concentração dos mesmos dados no ODS (Operational Data Source) ou *Staging Area*, que é um local de armazenamento temporário dos dados. O processo responsável que aperfeiçoa a criação da Data Warehouse e do Data Mart, possibilitando uma única fonte de dados sincronizada, é o ODS onde os dados originais são ajustados para a integração das informações. (RASLAN; CALAZANS, 2014)

Dentro do processo ODS, os dados acabam passando por etapas e sofrendo algumas transformações. As etapas são:

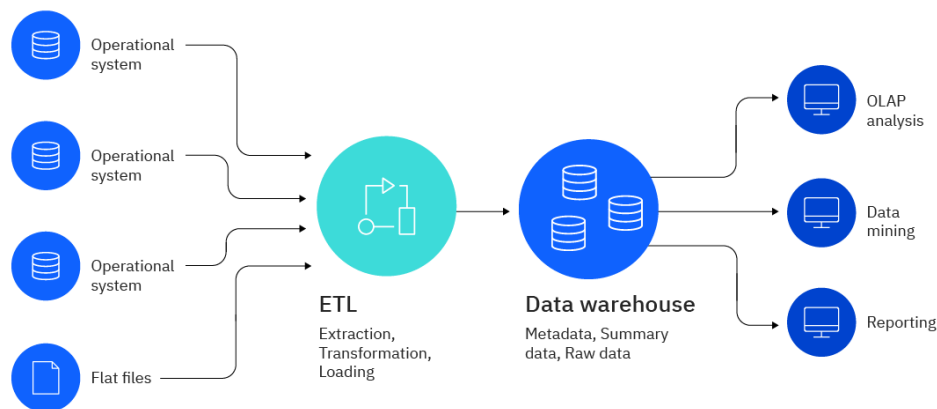
- **Integração:** criação de identificadores alternativos para cada registro, substituindo as chaves originais do sistema legado para evitar dependências diretas.
- **Limpeza:** processo de ajuste de dados, incluindo a remoção de códigos inválidos e caracteres especiais, solução de inconsistências nos domínios, tratamento de valores ausentes e correção de duplicadas de erros.
- **Eliminação:** exclusão de campos e informações herdadas dos sistemas legados que não têm utilidade no Data Warehouse (DW) ou Data Mart (DM).
- **Combinação:** fusão de dados com valores equivalentes para unificar informações dispersas.
- **Verificação de integridade referencial:** confirmação de que os dados relacionados entre diferentes tabelas que foram separadas durante a normalização em uma tabela desnormalizada.
- **Conversão de tipo de dados:** modificação do formato ou tipo dos dados para adequá-los a outro nível ou estrutura específica.
- **Cálculos, derivação e alocação:** aplicação de transformações baseadas nas regras de negócio identificadas durante a análise de requisitos.
- **Auditoria:** avaliação e validação do conteúdo dos dados para garantir sua qualidade e confiabilidade.

Esses meios utilizados para transformar e proceder a carga de dados, e assim recuperar, analisar e extrair os dados para gerenciar dados de dicionário são componentes essenciais para um Data Warehouse. (RASLAN; CALAZANS, 2014)

A arquitetura do DW consiste em três camadas principais: **Camada Inferior**, **Camada Intermediária** e **Camada Superior**. Cada uma desempenha

um papel específico no processamento e armazenamento de dados, assim como é possível observar na Figura 1. (IBM, 2024)

Figura 1



Fonte: IBM, 2024

A Camada Inferior pode ser resumida como um sistema de banco de dados relacional no qual tem como sua função em coletar, limpar e transformar dados de múltiplas fontes por meio de um processo conhecido como ETL (Extrair, Transformar e Carregar). Já a Camada Intermediária consiste em um servidor de processamento analítico online (OLAP) do qual possibilita velocidades de consulta rápidas. E por último a Camada Superior seria a representação de um tipo de interface de usuário ou até mesmo uma ferramenta de relatórios que permite aos usuários finais realizar as suas análises de dados em seus dados de negócios. (IBM, 2024).

A maior parte dos Data Warehouse (DW) são construídos ao redor de um sistema de banco de dados relacionais, podendo ser local ou em nuvem, na qual os dados são armazenados e processados. Um DW possui quatro elementos principais: o banco de dados central, as ferramentas ETL, metadados e as ferramentas de acesso. Todos esses componentes são projetados com a finalidade para velocidade, assim podendo obter os

resultados mais rápidos e analisar os dados em tempo real. Há também outros componentes que podem ser inclusos como um sistema de gerenciamento de metadados e uma camada de conectividade com a API, dessa forma permitindo que o armazém (repositório central) possa extrair os dados das fontes organizacionais. (IBM, 2024)

Como já foi dito antes, os DW foram projetados para armazenar grandes quantidades de dados chegando até a terabytes de informação, e às vezes, podem ser divididas em unidades menores chamadas de Data Mart (DM) (RASLAN; CALAZANS, 2014). O Data Mart representa um subconjunto de um Data Warehouse no qual é focado em uma determinada área de assunto ou departamento. Dessa forma disponibilizam dados específicos e um acesso rápido para um grupo definido de usuários, sem desperdiçar tempo em uma procura por todo Data Warehouse. (IBM, [s.d.])

Um Data Mart é criado a partir de um Data Warehouse existente por meio de outras tecnologias e ferramentas para projetar e construir um banco de dados físico por meio de um procedimento complexo. Existem três tipos de Data Marts (**Dependentes**, **Independentes** e **Híbrido**) que são diferentes em relação ao DW e as respectivas origens de dados de cada sistema. (IBM, [s.d.])

O Data Mart Dependente se trata de segmentos particionados de um DW corporativo, que no qual extraem um subconjunto a partir dos dados primários que são sempre necessários para análise. Em contrapartida, os Data Marts Independentes não necessitam de um DW para atuar, já que os analistas podem extrair os dados de um assunto específico, processá-los e armazená-los em um repositório de Data Mart que podem ser utilizados pela equipe. Por último, os Data Marts Híbridos combinam os dados de DW que já existem e de outras origens operacionais. (IBM, [s.d.])

2. Objetivos

Como foi apresentado anteriormente e com o contexto em que foi apresentado sobre a década que surgiu a tecnologia Data Warehouse, é possível afirmar que objetivo dessa década é para otimizar a análise de dados e gerenciar os dados. De acordo com JARKE et al. (2013), “Um Data

Warehouse (DW) é uma coleção de tecnologias destinadas a permitir que o trabalhador do conhecimento (executivo, gerente e analista) tome decisões melhores e mais rápidas. Espera-se ter as informações certas no lugar certo, na hora certa, com o custo certo, a fim de apoiar a decisão certa”.

Da mesma forma, os Data Marts que são unidades menores derivados do DW, também tem como objetivo o seu acesso mais rápido aos dados sem precisar realizar uma pesquisa por todo o Data Warehouse. Segundo JARKE et al. (2013), “Um data mart pode ser usado em um departamento específico e contém apenas os dados relevantes para esse departamento. Por exemplo, um data mart para o departamento de marketing deve incluir apenas informações sobre clientes, vendas e produtos, enquanto o Data Warehouse de toda a empresa também pode conter informações sobre funcionários, departamentos, etc”. Além disso, os Data Marts permitem respostas mais rápidas em relação as consultas por conta que o volume de dados gerenciados é muito menor comparado ao volume no DW.

3. Vantagens

Além das informações em relação ao Data Warehouse acerca da sua otimização a análise e gerenciamento de dados, ainda há outras vantagens sobre esta tecnologia. Os dados contidos em uma DW são inseridos com posições históricas das atividades e não se modificam, assim não podem ser excluídos ou modificados, apenas inseridos. (IBM, 2024)

A Data Warehouse oferece uma boa solução para as organizações com várias plataformas de hardware e software, que sofrem por falta de padronização e as várias alterações nos sistemas transacionais. Além de também oferecer um modelo de dados para diferentes áreas de interesse, facilitando reportar e analisar as informações. As inconsistências encontradas nas diferentes fontes de dados são resolvidas bem antes de serem carregadas e inseridas na DW, facilitando o processo de elaboração de relatórios e análise. (RASLAN; CALAZANS, 2014)

Um DW é capaz de centralizar dados de uma variedade de fontes de dados, podendo ser de sistemas transacionais, banco de dados operacionais e

arquivos planos. E, após a centralização dos diversos dados, o DW limpa os dados operacionais eliminando os dados duplicados e padronizando-os para poder criar apenas uma única fonte de verdade. Em contrapartida, o Data Mart se sobressai em sua eficiência de custos no momento da sua configuração utilizando o processo de ETL (extrair, transformar e carregar) e custando apenas uma fração do custo de uma Data Warehouse. (IBM, 2024)

Os Data Marts, como já ditos anteriormente, tem uma eficiência melhor em relação ao seu acesso aos dados por manterem um subconjunto pequeno de dados, assim os seus usuários podem recuperar os dados rapidamente com pouco trabalho em comparação ao acesso amplo de dados de um Data Warehouse. Além disso, a manutenção de dados de um Data Marts é muito mais fácil e é menos desorganizado por conta de focar em uma única linha e hospedam menos que 100GB. (IBM, 2024)

4. Desvantagens

Apesar de suas qualidades em relação à análise de dados, otimização, gerenciamento e a limpeza das duplicadas de informações, o Data Warehouse acaba sofrendo à medida que as empresas começam a armazenar mais dados e necessitam de mais análises avançadas, fora uma gama de dados. Dessa forma, RASLAN E CALAZANS (2014) cita que os dados podem acabar se tornando caro e não flexível, ficando obsoletos de forma muito rápida e causando prejuízos altos. (RASLAN; CALAZANS, 2014)

Além disso, os DW sofrem com o problema em relação para atender as demandas da gestão estratégica empresarial e a falta de apoio da alta gerência para o desenvolvimento de um Data Warehouse, segundo RASLAN e CAZALANS (2014). Outro detalhe é o extremo cuidado durante a sua modelagem para que possa atender as expectativas de seus patronos, fornecendo as informações corretas e estratégicas, caso contrário acaba-se tornando apenas repetições dos sistemas transacionais que existem na corporação. (RASLAN; CALAZANS, 2014)

As formas de como os dados são armazenados depois de serem recuperados, analisados, extraídos e gerenciados podem trazer também as

suas desvantagens dependendo da escolha. Em um Data Warehouse, possui duas abordagens para o armazenamento: abordagem normalizada e abordagem dimensional.

A abordagem normalizada trata-se dos dados armazenados seguindo a regra de Codd, na qual as tabelas são agrupadas por temas nas quais refletem as categorias dos dados gerais. Sua desvantagem está presente em sua dificuldade para que os usuários possam unir os dados que vieram de diferentes fontes de informações. (RASLAN; CALAZANS, 2014)

Na abordagem dimensional, os dados são divididos em fatos e dimensões. Fatos são aqueles dados numéricos de transações e as dimensões se trata das informações de referência que contextualizam os fatos. E a sua desvantagem aparece por conta da sua complexidade no carregamento de dados por meio dos diferentes sistemas operacionais e na dificuldade de modificar a sua estrutura de armazenamento de dados. (RASLAN; CALAZANS, 2014)

De acordo com o estudo de caso feito por RASLAN e CALAZANS (2014) com intuito de aplicar um questionário a oito profissionais de TI, que trabalharam em empresas diferentes, a fim de verificar se havia correspondência com os conteúdos que foram estudados e obtidos em relação ao DW e o uso cotidiano dos conceitos pelos profissionais na área, trouxe informações em relação as desvantagens que encontraram na adoção da arquitetura de Data Warehouse. Os profissionais entrevistados, que não relataram ter sido vantajosa a adoção de DW, relataram que o principal impedimento foi por conta da falta de preparo e a maturidade da organização para aplicar a tecnologia, além da falta de integração de dados, questões políticas e a gestão sem objetividade.

5. Exemplos de uso

É possível notar que o uso de Data Warehouse (DW) e Data Mart (DM) está crescendo ao redor das diversas indústrias por conta da sua capacidade de centralizar e otimizar a análise de grandes volumes de dados, como por exemplo o estudo de caso publicado pela IBM (2025) em que o produto DB2

Warehouse oferecido pela IBM e que é utilizado pela empresa Active Internacional para que possa desenvolver as aplicações com o intuito de otimizar o ciclo de compra de mídia e auxiliar na identificação de novas perspectivas de negócios.

Outro grande sucesso com o uso de Data Warehouse foi com a Capital Bank da Jordânia, que oferece serviços e soluções bancárias comerciais e de investimentos para os clientes corporativos. De acordo com o estudo de caso publicado pela IBM (2025), a empresa fez a adoção da tecnologia Data Warehouse, do produto Netezza Performance Server da IBM, que trouxe impactos positivos significativos que reduziu o tempo de migração de dados relacionados à aquisição em mais de 95%, assim permitindo uma integração mais rápida das operações.

Apesar dos benefícios e das vantagens citadas anteriormente, a tecnologia e os processos de adesão de Data Warehouse ainda estão em processo de crescimento em vista das dificuldades que são encontradas para integrar com outros sistemas e, caso mal aplicado, possa gerar inconsistência de dados redundantes, causando um prejuízo para a empresa que o adotou.

6. Base de Dados

O nome da base de dados, escolhida para servir como estudo de caso em relação a pesquisa acerca de Data Warehouse, se chama *SMMnet – Super Mario Maker*. A base de dados contém enormes dados em relação aos 115 mil mapas criados no jogo Super Mario Maker e com a participação de mais de 880 mil jogadores que realizaram mais de 7 milhões de interações nesses mapas.

A estrutura da base de dados contém 7 tabelas que foram organizadas para capturar os dados relacionados aos jogadores, mapas e alterações temporais. Foi projetado com o intuito de integração em sistemas de gerenciamento de banco de dados relacionais, facilitando a associação entre mapas e jogadores por meio de chaves primárias e estrangeiras.

A documentação a seguir se tratam, respectivamente, do DE-R (Diagrama de Entidade Relacionamento) e DLD (Diagrama Lógico de Dados)

da base de dados *SMMnet – Super Mario Maker* para se ter uma melhor visão e entendimento da base de dados e os seus relacionamentos.

Figura 1 – DE-R

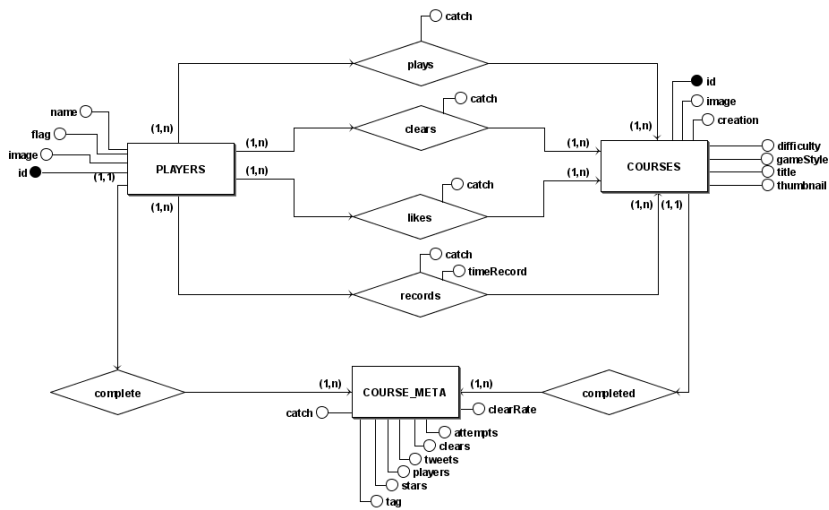
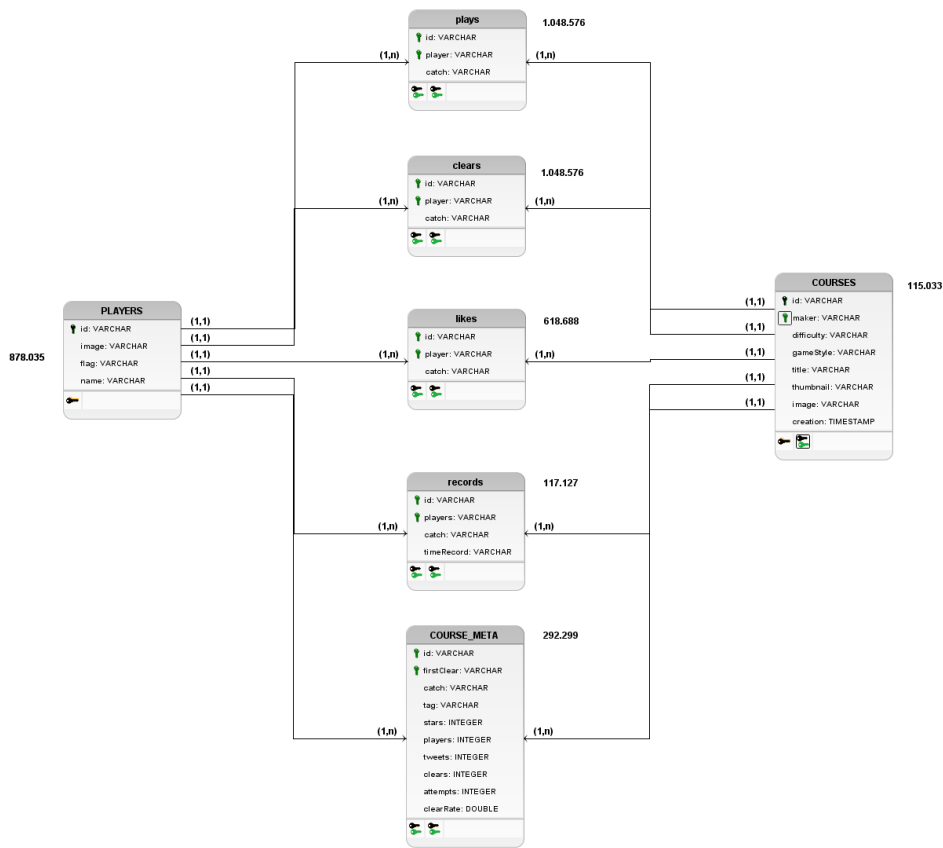


Figura 2 - DLD



Endereço Virtual da Base de Dados:
<https://www.kaggle.com/datasets/leomauro/smmnet?select=clears.csv>

7. Referências Bibliográficas

JARKE, M. ET AL. **FUNDAMENTALS OF DATA WAREHOUSES**. 2. ED. BERLIM, GERMANY: SPRINGER, 2013.

O QUE É DATA WAREHOUSE? IBM.COM, 16 AGO. 2024. DISPONÍVEL EM: <[HTTPS://WWW.IBM.COM/BR-PT/TOPICS/DATA-WAREHOUSE](https://www.ibm.com/br-pt/topics/data-warehouse)>. ACESSO EM: 18 JAN. 2025

O QUE É UM DATA MART. DISPONÍVEL EM: <[HTTPS://WWW.IBM.COM/BR-PT/TOPICS/DATA-MART](https://www.ibm.com/br-pt/topics/data-mart)>. ACESSO EM: 18 JAN. 2025.

RASLAN, D. A.; CALAZANS, A. T. S. **DATA WAREHOUSE: CONCEITOS E APLICAÇÕES** - DOI: 10.5102/UN.GTI.V4I1.2612. UNIVERSITAS GESTÃO E TI, V. 4, N. 1, 2014.

CAPITAL BANK OF JORDAN. DISPONÍVEL EM: <[HTTPS://WWW.IBM.COM/BR-PT/CASE-STUDIES/CAPITAL-BANK](https://www.ibm.com/br-pt/case-studies/capital-bank)>. ACESSO EM: 19 JAN. 2025.

IBM DB2 WAREHOUSE. DISPONÍVEL EM: <[HTTPS://WWW.IBM.COM/BR-PT/PRODUCTS/DB2-WAREHOUSE](https://www.ibm.com/br-pt/products/db2-warehouse)>. ACESSO EM: 19 JAN. 2025.