

Mestrado em Engenharia Informática

Perfis de Especialização, Edição 2024-2025

Armazéns de Dados

Sistemas de Data Warehousing

Orlando Belo

ALGORITMI / LASI

Departamento de Informática, Escola de Engenharia, Universidade do Minho

Novembro, 2024





SINGLE SOURCE OF TRUTH

In information science and information technology, single source of truth architecture, or single point of truth architecture, for information **systems is the practice of structuring information models and associated data schemas such that every data element is mastered in only one place**, providing data normalization to a canonical form. Wikipedia



Estrutura

- Contextualização
- Fundamentação
- Objetivos
- Plano Curricular
- Requisitos
- Referências



Contextualização

- A implementação de um sistema de data warehousing no seio de uma organização disponibiliza um conjunto de meios altamente eficientes para o armazenamento consistente de dados sem ruído, organizados segundo os diversos pontos de decisão e dimensões de análise dos agentes empresariais, bem como serviços adequados para a manipulação e exploração desses dados de forma simples e flexível.

SINGLE SOURCE OF TRUTH



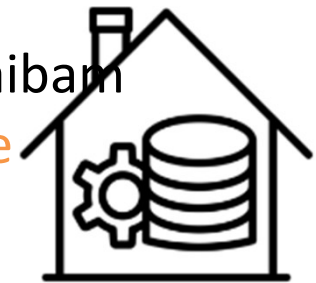
Fundamentação

- O perfil de especialização em Sistemas de Data Warehousing pretende transmitir aos alunos o conhecimento necessário para:
 - planearem, justificarem, desenharem, implementarem e gerirem um **projeto de um sistema de data warehousing**
 - serem capazes de **explorarem** de forma efetiva os **data warehouses** desenvolvidos, especificamente para suporte aos processos de decisão empresariais que regularam a definição das suas estruturas multidimensionais de dados.



Objetivos

- Estudar as principais **técnicas de planeamento estratégico** para a integração de um sistema de data warehousing no seio de uma organização
- Fornecer o conhecimento necessário para a **definição, gestão e manutenção** de um projeto de implementação de um **sistema de suporte à decisão** no seio de uma organização.
- Providenciar as bases necessárias para o desenvolvimento de processos de **análise, planeamento, desenho e conceção de modelos para data warehouses**.
- Transmitir o conhecimento essencial para que os alunos saibam administrar e aplicar **modelos de segurança, recuperação e privacidade de dados** em sistemas de data warehousing.



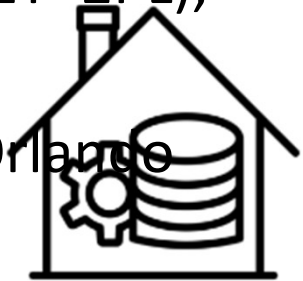
Objetivos

- Estudar ferramentas de **modelação e desenvolvimento de sistemas de data warehousing** atualmente disponíveis no mercado.
- Analisar e aplicar modelos para a **migração de dados entre diferentes sistemas de dados** e povoamento de sistemas de data warehousing.
- Aprender a **explorar os data warehouses** e as estruturas de dados multidimensionais através de técnicas de extração de conhecimento e de análise de dados adequadas.



Plano Curricular

- O perfil de especialização em Sistemas de Data Warehousing será lecionado durante o 2º semestre e integra no seu plano curricular as seguintes unidades de ensino:
 - **Conceção e Implementação de Data Warehouses (1T+2PL)**, Orlando Belo.
 - **Sistemas de Extração, Transformação e Carregamento de Dados (1T+2PL)**, António Abelha.
 - **Extração de Conhecimento em Data Warehouses (1T+2PL)**, César Analide.
 - **Projeto em Sistemas de Data Warehousing (2PL)**, Orlando Belo.



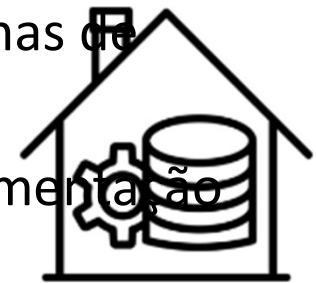
Conceção e Implementação de Data Warehouses

- Introdução aos **Sistemas de Data Warehousing**.
- **Arquiteturas e Serviços** de um Sistema de Data Warehousing.
- **Ciclo de vida do desenvolvimento de um Data Warehouse**.
- **Projeto** de um Sistema de Data Warehousing.
- **Modelação Dimensional** de Dados.
- Topologias e organização de **esquemas multidimensionais de dados**.
- Estratégias e sistemas de povoamento de Data Warehouses.
- Plataformas analíticas para a exploração de Data Warehouses.
- Utilização de **Data Vaults** em Sistemas de Data Warehousing.
- Implementação de **Data Lakehouses**.



Sistemas de Extração, Transformação e Carregamento de Dados

- Introdução aos **sistemas de povoamento** de data warehouses.
- **Projeto e implementação** de sistemas de povoamento de data warehouses.
- Modelos e serviços para **extração, transformação e carregamento de dados** em data warehouses.
- Conceção e implementação de **áreas de estágio** para sistemas de povoamento.
- **Workflowing** de processos para o povoamento de data warehouses.
- Modelos e serviços de recuperação de processos de sistemas de povoamento.
- Aplicações e ferramentas para o desenvolvimento e implementação de sistemas de povoamento de data warehouses.



Extração de Conhecimento em Data Warehouses

- Introdução aos **sistemas de aquisição de conhecimento**.
- Ciclo de vida da descoberta de conhecimento.
- Introdução à **aprendizagem máquina** e **mineração de dados**.
- Preparação e pré processamento de dados.
- Linguagens e arquiteturas para mineração de dados em data warehouses.
- Técnicas e modelos de **classificação e de predição** de dados.
- Mineração de regras de **associação** e de sequências.
- Geração e análise de **clusters**.
- Mineração de tipos de dados complexos.

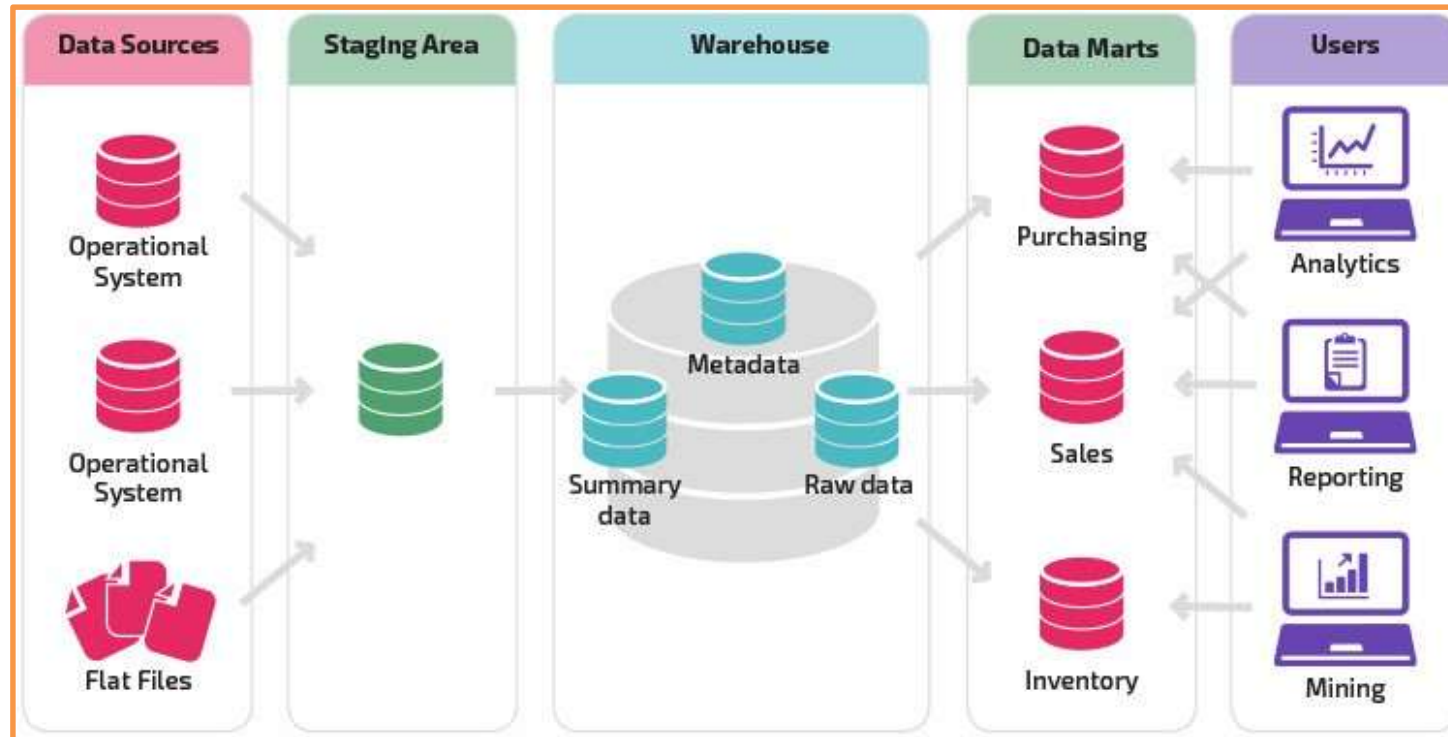


Projeto em Sistemas de Data Warehousing

- **Engenharia** de Sistemas de Sistemas de Data Warehousing.
- **Metodologias** para o Desenvolvimento de Sistema de Data Warehousing.
- **Projeto**, Implementação, Validação e Testes de Sistemas de Data Warehousing.
- Software para a **Modelação** e **Desenvolvimento** de Sistemas de Data Warehousing.



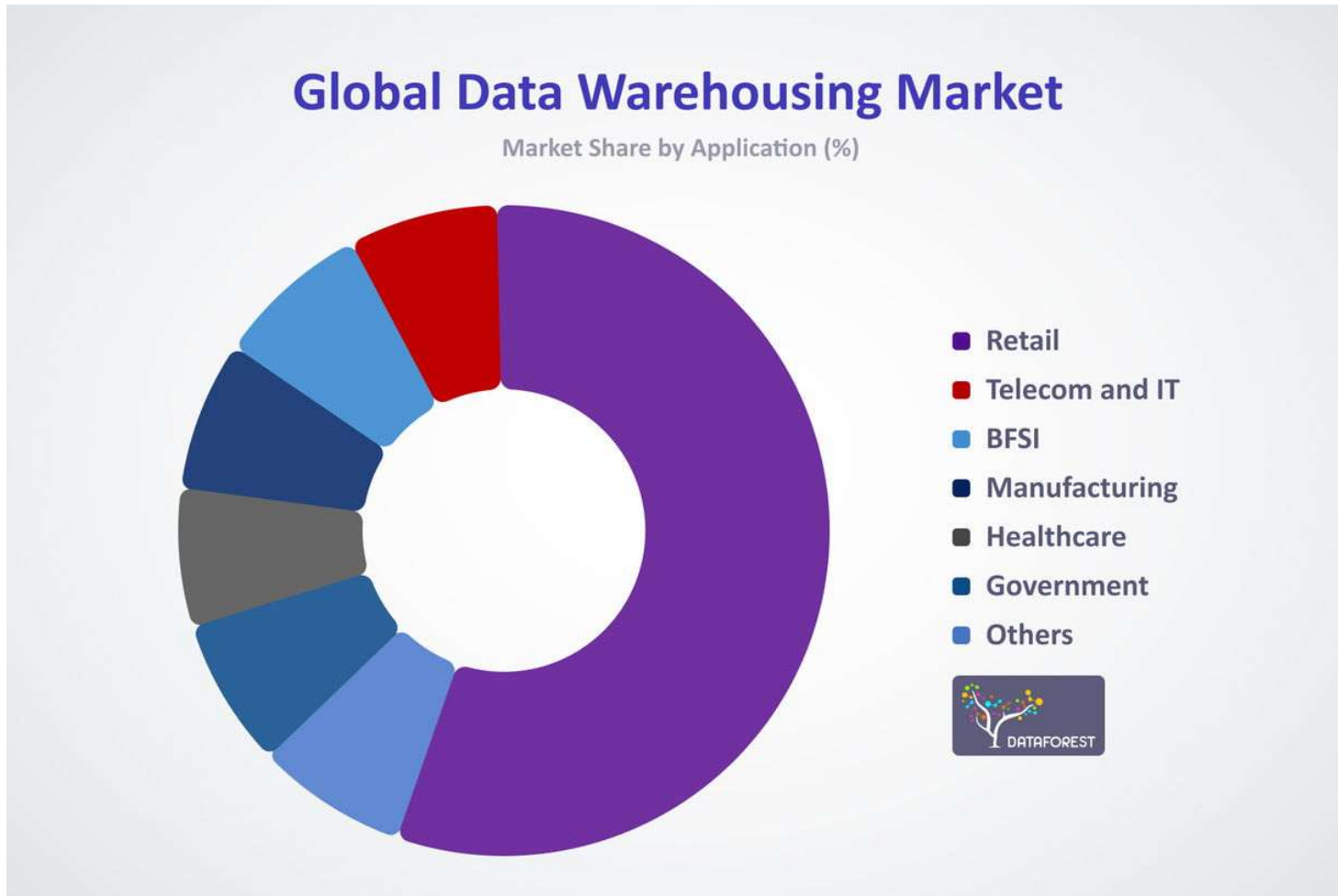
Um Sistema de Data Warehousing



Source: <https://www.astera.com/knowledge-center/data-warehouse-architecture/>



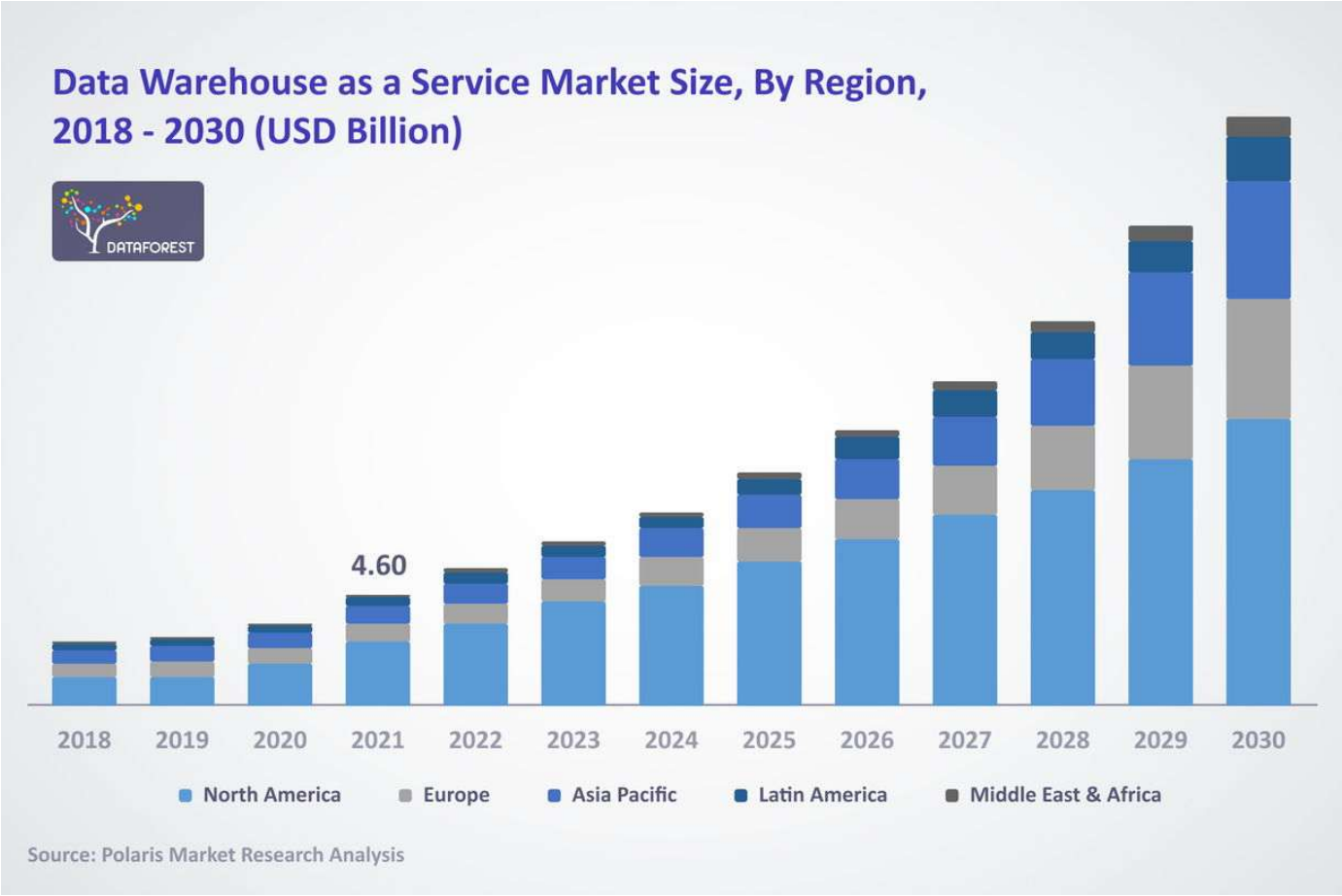
O Mercado



Source: <https://dataforest.ai/blog/practical-data-warehousing-successful-cases>



Evolução



Source: <https://dataforest.ai/blog/practical-data-warehousing-successful-cases>



Exemplos de Data Warehouses

- **Walmart** - uma das maiores implementações de data warehousing do mundo.
- **Amazon** - a solução de armazenamento de dados da Amazon é conhecida como Amazon Redshift.
- **Netflix** - para armazenar e analisar dados da sua plataforma de streaming.
- **Coca-Cola** - para consolidar dados de diferentes unidades de negócios e analisá-los.
- **Bank of America** - para analisar dados de clientes com o objetivo de melhorar a experiência do cliente.





Source: https://www.youtube.com/watch?v=AHR_7jFCMeY



(Pre)Requisitos

- Para frequentar este perfil de especialização não são requeridos quaisquer requisitos particulares. Apenas se espera que os alunos candidatos:
 - tenham **conhecimento e perícia** adquiridas no domínio dos **sistemas de bases de dados**;
 - (achem que) gostam das áreas da **engenharia de dados e análise de dados**;
 - reconheçam a importância e o relevo da **aplicação do conhecimento** adquirido nas áreas referidas em **aplicações do mundo real**.



Mestrado em Engenharia Informática

Perfis de Especialização, Edição 2024-2025

Sistemas de Data Warehousing

Orlando Belo

ALGORITMI / LASI

Departamento de Informática, Escola de Engenharia, Universidade do Minho

Novembro, 2024

